DISSERTATION

QUANTIFYING INTERNAL CLIMATE VARIABILITY AND ITS CHANGES USING LARGE-ENSEMBLES OF CLIMATE CHANGE SIMULATIONS

Submitted by

Jingyuan Li

Department of Atmospheric Science

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2020

Doctoral Committee:

Advisor: David W. J. Thompson

Elizabeth A. Barnes
A. R. Ravishankara
Daniel Cooley

ABSTRACT

QUANTIFYING INTERNAL CLIMATE VARIABILITY AND ITS CHANGES USING LARGE-ENSEMBLES OF CLIMATE CHANGE SIMULATIONS

Increasing temperatures over the last 50 years have led to a multitude of studies on observed and future impacts on surface climate. However, any changes on the mean need to be placed in the context of its variability to be understood and quantified. This allows us to: 1) understand the relative impact of the mean change on the subsequent environment, and 2) detect and attribute the external change from the underlying “noise" of internal variability. One way to quantify internal variability is through the use of large ensemble models. Each ensemble member is run on the same model and with the same external forcings, but with slight differences in the initial conditions. Differences between ensemble members are due solely to internal variability. This research exploits one such large ensemble of climate change simulations (CESM-LE) to better understand and evaluate surface temperature variability and its effects under external forcing.

One large contribution to monthly and annual surface temperature variability is the atmospheric circulation, especially in the extratropics. Dynamical adjustment seeks to determine and remove the effects of circulation on temperature variability in order to narrow the range of uncertainty in the temperature response. The first part of this work compares several commonly used dynamical adjustment methods in both a pre-industrial control run and the CESM-LE. Because there are no external forcings in the control run, it is used to provide a quantitative metric by which the methods are evaluated. We compare and assess these dynamical adjustment methods on the basis of 2 attributes: 1) the method should remove a maximum amount of internal variability while 2) preserving the true forced signal. While the control run is excellent for assessing the methods in an “ideal” environment, results from the CESM-LE show biases in the dynamically-adjusted trends

due to a forced response in the circulation fields themselves. This work provides a template from which to assess the various dynamical adjustment methods available to the community.

A less studied question is how internal variability itself will respond to climate change. Past studies have found regional changes in surface temperature variance and skewness. This research also investigates the impacts of climate change on day-to-day persistence of surface temperature. Results from the CESM-LE suggest that external warming generally increases surface temperature persistence, with the largest changes over the Arctic and ocean regions. The results are robust and distinct from internal variability. We suggest that persistence changes are mostly due to an increase in the optical thickness of the atmosphere due to increases in both carbon dioxide and water vapor. This increased optical thickness reduces the thermal damping of surface temperatures, increasing their persistence. Model results from idealized aquaplanet simulations with different radiation schemes support this hypothesis. The results thus reflect a robust thermodynamic and radiative constraint on surface temperature variability.

ACKNOWLEDGMENTS

This dissertation would not have been possible without the support of many others. First, my advisor, Dave Thompson, for over 6 years of support, guidance, and wisdom. I would also like to thank my committee - Libby Barnes, Ravi, and Dan Cooley, for their time serving on my committee and providing invaluable feedback. Special thanks to the students and staff in the Thompson research group, both past and present. Finally, I'd like to thank my family and friends for their continuous support, love, and encouragement during their entire process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 Introduction

It is evident that the global-mean surface air temperature (SAT) has increased throughout the 20th and 21st centuries, with an increase of around 0.85°C over the period 1880 to 2012 (Hartmann et al. (2013)). However, this warming has not been temporally uniform. For instance, the so-called “hiatus" period from 1998-2013 resulted in a period with almost no change in the global mean surface temperatures, leading to a flurry of publications on the topic (e.g., Lewandowsky et al. (2016); Yan et al. (2016)). Nor is the change spatially uniform across the globe. The Arctic has been warming at twice the rate of the Northern Hemisphere land masses over the past 50 years (Walsh (2014)), and land temperatures have warmed much more than sea surface temperatures (e.g., Sutton et al. (2007)).

This study seeks to better understand the variability in surface temperature, its changes under global warming, and possible impacts in future climate. It examines the role of circulation patterns on SAT variability and investigates changes in SAT persistence under SAT warming. This chapter provides a brief literature review of the problem of internal variability in climate projections, SAT variability and its response to anthropogenic warming, the detection and attribution of climate change, and provides an overview of large ensemble models. It concludes with an overview of the remaining chapters in this dissertation.

1.1 INTRODUCTION TO INTERNAL VARIABILITY

In future climate projections, uncertainty arises from three main sources: 1) uncertainty in anthropogenic emissions, 2) uncertainty in the model response (differences amongst climate models), and 3) uncertainty from the internal variability of the climate system (Hawkins and Sutton (2009)). Internal variability is defined as the natural (innate) variability of the climate system arising from processes in and coupled interactions between the atmosphere, oceans, land, and cryosphere (e.g., Schneider and Kinter 1994; Deser et al. 2012b; Kay et al. 2015). It manifests itself on a multitude

of timescales, from small atmospheric waves on the order of seconds to minutes to multidecadal patterns of variability such as the Atlantic Multidecadal Oscillation (AMO).

While we can better specify emissions scenarios and improve our climate models, the random processes in the unforced climate system makes it difficult to reduce the uncertainty from internal variability. As such, internal variability poses an inherent limit to climate predictability. Although human-induced climate change is likely to dominate over internal variability for time periods longer than a decade on the global scale, internal variability can overwhelm anthropogenic forcing on regional and local scales, even on timescales of up to 50 years at middle and high latitudes (Deser et al. (2012b), Deser et al. (2014)). At the same time, opposing signs of internal variability may mask or hide the effects of anthropogenic forcing. Quantifying uncertainty is especially important for climate risk assessments, e.g., in the health, agriculture, energy, traffic or infrastructure sectors. The understanding of the internal climate variability has been identified as one of the hardest geophysical problems of the twenty-first century (Ghil (2001)).

Many previous studies have shown the significant role internal variability plays in estimating climate change for both the historical record and for future projections of climate change. For instance, Deser et al. (2012a) found trend differences of up to 6°C in projections of temperature over North America for the period 2005-2060 due to internal variability alone. Precipitation was found to have an even wider spread than temperature. This makes it difficult to interpret future climate projections, especially when multiple models are used (such as when assessing the CMIP3 or CMIP5 model means). Indeed, Deser et al. (2012b) suggested that at least half of the inter-model spread amongst CMIP3 models for the period 2005-2060 is due to internal variability alone.

1.2 SAT VARIABILITY AND CHANGE

This dissertation primarily focuses on variability of surface temperatures. SAT (along with SST) is a commonly used metric for climate monitoring. SAT variability is generally thought of as a response to forcings arising from dynamic, thermodynamic, and radiative processes in the oceans and atmosphere (Compo and Sardeshmukh (2007)). As an example, Figure 1.1 compares land and ocean global-mean temperature anomalies over the latter half of the 20th century. Land

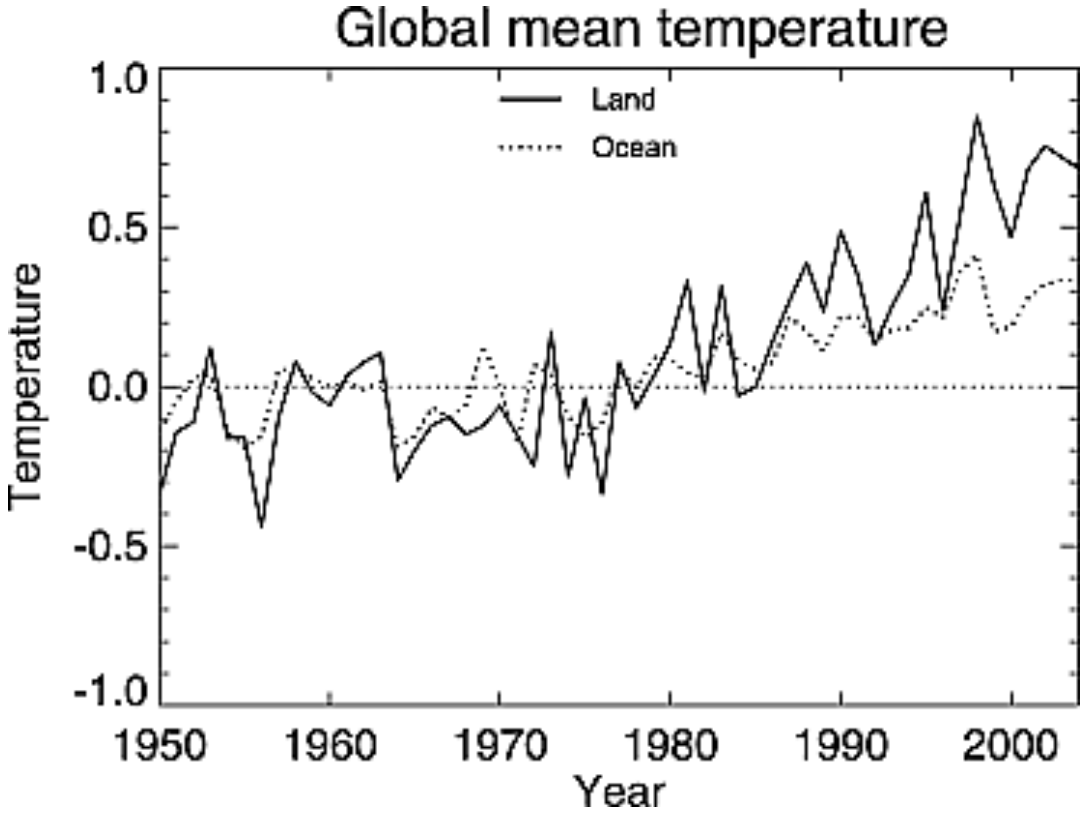

Fig. 1.1. Reprinted from Sutton et al. (2007). Global mean temperature anomalies in observations for land vs. ocean regions.

temperatures exhibit greater variability than ocean temperatures, due to the larger heat capacity of the oceans. The larger global warming trend over land relative to ocean has been attributed to differences in their surface energy budgets (Sutton et al. (2007)).

While many studies have focused on quantifying internal variability in the context of global warming, several studies have also looked at how internal variability itself may respond to climate change. One of the greatest impacts of a changing variability is the effect on extreme weather events (i.e., heat waves, extreme cold events, flooding, droughts, etc.). While there has been significant interest in examining the influence of global warming on extreme events, studies have shown that the variability may matter more than the mean (Katz and Brown (1992)). Figure 1.2 comes from the IPCC AR5 report and shows an example of how extreme events can change under changes in the mean, variance, and/or skewness of the underlying distribution.

Recent heat waves have sparked interest in examining changes in temperature variability. While some studies suggest that these events were caused in part by a broadened temperature

distribution and increased temperature variability (e.g., Schär et al. (2004a), Hansen et al. (2012)), other studies disagree, arguing that current temperature extremes are consistent with a simple shift of the temperature distribution and not due to any changes in higher moments (e.g., Rhines and Huybers (2012), Huntingford et al. (2013), Simolo et al. (2011), Rahmstorf and Coumou (2011), Coumou and Robinson (2013)).

While temperature variability changes are inconclusive in observations, modeling studies for the mid to late 21st century have found several robust changes under a variety of warming scenarios (e.g., Räisänen (2002), Kharin et al. (2007), Boer (2009), Huntingford et al. (2013), LaJoie and DelSole (2016)). Overall, these studies suggest that temperature variability decreases over mid and high latitudes, and increases slightly over the tropics. There is also a strong seasonality – changes in the high latitudes are much greater in winter, while in summer the midlatitudes experience an increase in variance (versus a decrease in winter and in the annual mean) (e.g., Stouffer and Wetherald (2007), Scherrer et al. (2008), Holmes et al. (2016)). LaJoie and DelSole (2016) found that, while changes to internal variability in any single component are not large enough to be significant, the surface temperature internal variability changes are field significant overall.

A decrease in temperature variability over the Arctic region due to sea ice loss has been shown to be a robust response to climate change (e.g., Räisänen (2002), Boer (2009), Screen (2014), LaJoie and DelSole (2016), Yettella et al. (2018)). The increase in summer temperature variability is also fairly consistent across models and studies, and has been linked to soil moisture-temperature feedbacks (e.g., Screen (2014), Fischer et al. (2012)). Increased temperatures lead to a drying of continental land due to a depletion of soil moisture, which in turn affect the partitioning of latent and sensible heat fluxes controlled by the soil moisture (e.g., Stouffer and Wetherald (2007), Scherrer et al. (2008), Fischer et al. (2012), Yettella et al. (2018)).

Changes in circulation have also been found to impact both temperature and precipitation variability. This uneven warming rate of Arctic Amplification reduces the meridional temperature gradient between the Arctic and midlatitudes, leading to changes in thermal advection (e.g., Schneider et al. (2015), Holmes et al. (2016)). Temperature in the mid and high latitudes have been found to be correlated to wind direction (cold days are associated with northerly winds, while warm days

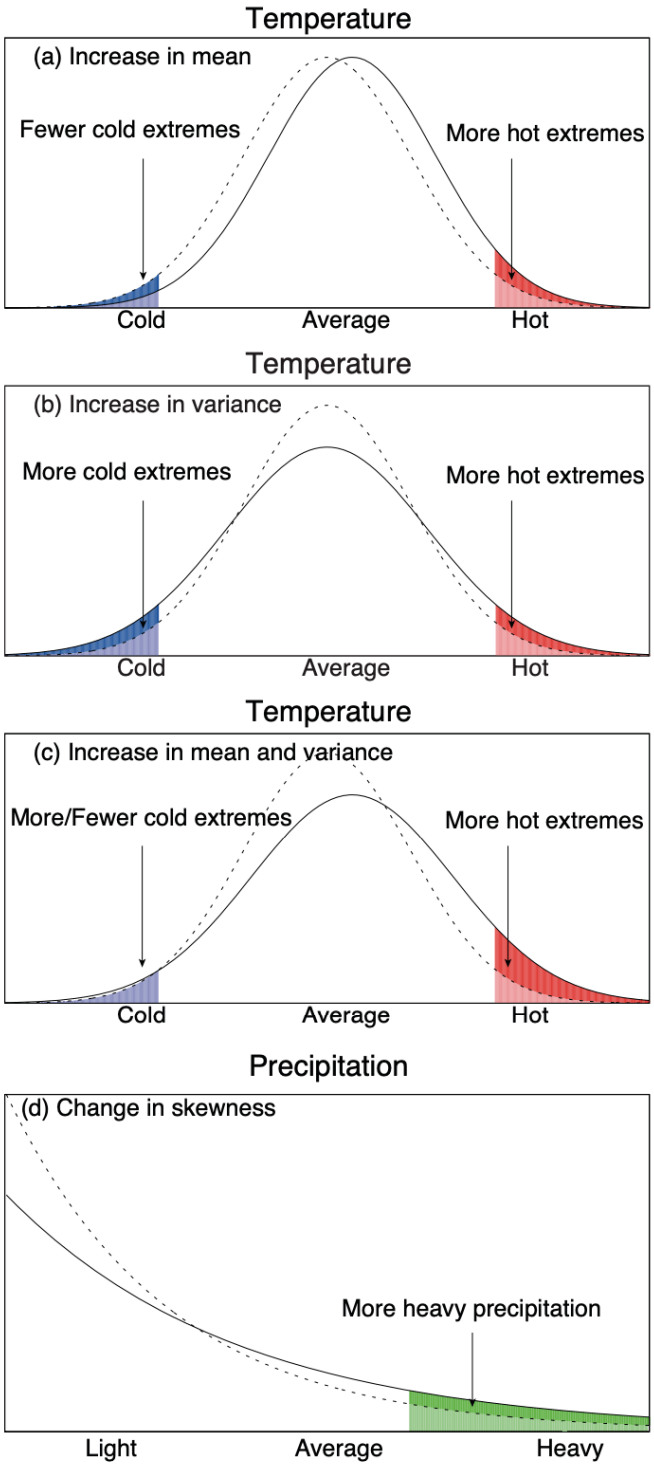

Fig. 1.2. Reprinted from Hartmann et al. (2013). A schematic showing extreme events for (a) an increase in the mean temperature; (b) and increase in the variance; (c) an increase in both the mean and the variance; and (d) a change in the skewness of precipitation.

are associated with southerly winds), and projected decreases in temperature variability have been linked to changes in heat advection due to Arctic amplification (e.g., Screen (2014)). Circulation anomalies and advective effects associated with the response to anthropogenic warming contribute to changes in regional frost days and precipitation over the midlatitudes (Meehl et al. (2004)(Meehl et al. (2004), Meehl et al. (2005), Kooperman et al. (2015)). Screen (2013) find that summer precipitation over northern Europe is associated with the position of the jet stream; more precipitation occurs when the jet stream is displaced to the south, and vice versa.

Higher-order moments of the temperature distribution have also been examined, albeit with less attention. Both Garfinkel and Harnik (2017) and Tamarin-Brodsky et al. (2018) examine the connections between dynamical processes in the midlatitudes with extreme temperature events. They find an asymmetrical and non-Gaussian distribution of surface temperature variability due advection by the storm tracks. These studies suggest that skewness, along with variance, is important for understanding and projecting future temperature variability under climate change.

1.3 AN OVERVIEW OF LARGE ENSEMBLE MODELS

One way to quantify the intrinsic internal variability in the climate system is by use of an ensemble of climate simulations. A single coupled climate model with a specified radiative forcing scenario is run many times with slightly different initial conditions, generating an “ensemble" of different realizations. After the memory of the initial conditions is lost, each member evolves independently from internal fluctuations characteristic of a random, stochastic process (Deser et al. (2012b)). Initial conditions are created either from 1) a “micro" perturbation, in which a very small perturbation (level of round-off error) is applied into the atmosphere of each ensemble member with all else held equal; or 2) a “macro" perturbation, in which all ensemble members are started with different initial conditions in the ocean and other components. The “micro" initial condition is best for studies involving atmospheric variability and trends, while the “macro" initial condition is best for studies on longer-scale ocean and climate variability (Deser et al. (2020)).

Large ensemble models provide an accurate way to quantify the internal variability in each individual model. Furthermore, by averaging the individual ensemble members, we can also obtain the forced response of the climate model to the imposed external forcings. This is in contrast to observations, where both the forced response and the internal variability have to be estimated (commonly as the variance and linear trend of the signal) and impossible to separate. The known internal variability and forced response from large ensembles also allow for detection and attribution studies, the point when the “signal" of climate change can be detected from the “noise" of internal variability (e.g., Li et al. (2017)).

Increases in technological and computational power have led to an explosion in the number and size of large ensemble models. While large ensembles have rapidly been gaining popularity in the climate sciences, they are not without downsides. Each single large ensemble requires substantial high-performance computing resources to run, and large amounts of memory space in which to store outputs. Making these datasets accessible for the wider audience is also challenging. There are also many questions on the models' representation of internal variability. If the large ensemble model does not represent the internal variability of the Earth's climate system, is it still useful? Several studies have sought to address this issue; for instance, an “observational large ensemble" created from statistical resampling methods applied to observations has been used to assess the internal variability simulated by the NCAR CESM1 Large Ensemble model (McKinnon et al. (2017)). Regardless, this remains an important challenge for the large ensemble community.

1.4 OVERVIEW OF DISSERTATION

This dissertation aims to provide new insights into surface temperature variability and changes in future climate. While this work utilizes a large ensemble model (the NCAR CESM-LE), our research does not focus explicitly on quantifying or assessing the model's representation of internal variability. Rather, the unique framework provided by the large ensemble allows us to examine specific aspects of temperature variability in more detail. The chapters are organized as follows. Chapter 2 provides a description of both the data and methods used throughout the dissertation. Chapters 3 and 4 are research chapters, each written as a stand-alone paper. The results in Chapter

3 will be submitted to a peer-reviewed journal in the coming months, while Chapter 4 has recently been submitted, and is currently under the review process.

In Chapter 3, we provide a comprehensive analysis and comparison between common methods used for dynamical adjustment (identifying the role of internal dynamics on temperature trends). The methods are compared in both a pre-industrial control run and the large ensemble simulations from the CESM-LE. We test each method quantitatively using the root-mean-square-error metric. We also examine the methods' biases under external forcing and discuss the advantages and disadvantages of different methods. Chapter 4 explores the response of temperature persistence on subseasonal timescales to external climate forcing in the CESM-LE. Surface temperature persistence generally increases in a warmer climate, with the most pronounced changes over the oceans and the Arctic. These changes are most likely due to 1) an increase in surface heat capacity over the Arctic, and 2) increased optical thickness of the atmosphere due to increased water vapor and greenhouse gases. Lastly, Chapter 5 summarizes and discusses the main results of this dissertation. It also suggests several avenues for future research.

2 Data and Methods

2.1 DATA

2.1.1 NCAR CESM1 model

This dissertation makes use of both a preindustrial control run and a large ensemble of climate change simulations from the National Center for Atmospheric Research (NCAR) CESM Large Ensemble Project. The models are run on a nominal 1 degree latitude/longitude resolution from a single CMIP5 coupled climate model: the Community Earth System Model version 1 (CESM1), with the Community Atmosphere Model, Version 5.2 [CESM1(CAM5.2)] as its atmospheric component. The model is fully-coupled, consisting of atmosphere, ocean, land, and sea ice prognostic component models (Kay et al. (2015)). The data was obtained from NCAR's Climate Data Gateway (http://www.cesm.ucar.edu/projects/community-projects/LENS/data-sets.html)

The control simulations are run for 1800 years under pre-industrial (1850) external forcings. The large ensemble simulations (CESM-LE) consist of 40 ensemble members for the period 1920-2100. The first 35 ensemble members were run at NCAR, while the latter ensemble members were run at the University of Toronto. Each ensemble member was run with the exact same external forcings (historical until 2005 and RCP8.5 thereafter) but begin from slightly different initial conditions (Kay et al. (2015)). As such, after an initial spin-up period (on the order of weeks in the atmosphere), all differences between the ensemble members are due to internal variability. An overview of the global-mean surface temperatures anomalies from the CESM control run, all ensemble members in the CESM-LE, and observations can be seen in Figure 2.1. The internal variability can be seen as the spread amongst the 40 ensemble members, and is shown against the observational record. The temperature trend from 1850 to 2100 is also shown. Additional details of the simulations can be found in Hurrell et al. (2013) and Kay et al. (2015). Key variables from the model runs used in this dissertation include: surface air temperature, sea level pressure (SLP), zonal wind at 950hPa (U950), and meridional wind at 950hPa (V950).

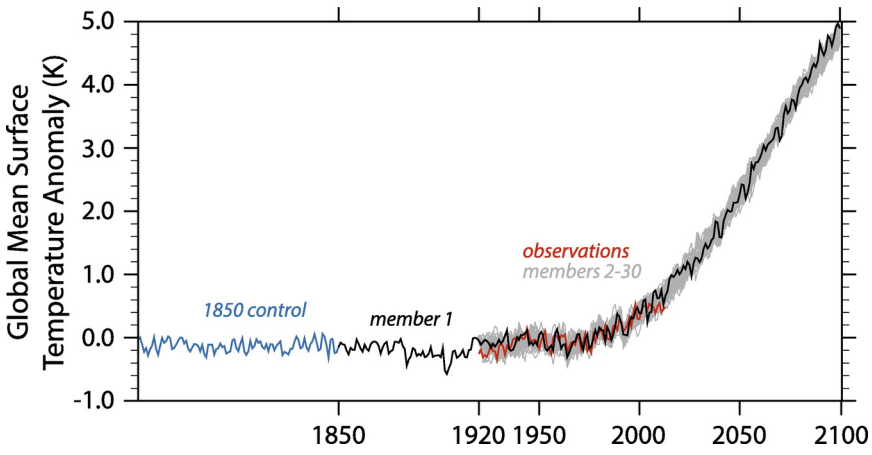

Fig. 2.1. Reprinted from Kay et al. (2015). Global surface temperature anomalies (relative to the 1961-90 period) for the pre-industrial control run, all ensemble members, and observations from HadCRUT4.

2.1.2 Aquaplanet simulations

Chapter 4 also makes use of several aquaplanet simulations run using two different radiation schemes that were part of the “longwave hierarchy" of the numerical model configurations in Tan et al. (2019). An aquaplanet is defined as an idealized configuration in which a planet’s surface is completely covered by water, and as such does not include any land, topography effects, or sea ice. Following Kang et al. (2008), the model consists of a slab-ocean with a mixed layer depth of 30m, and thus also excludes ocean energy transport. The atmospheric component is from the GFDL AM2.1 model (Delworth et al. (2006)) and is run at a uniform horizontal resolution of 2.8° latitude by longitude, 48 vertical levels, and a time step of 900 seconds.

Two radiation schemes are used: the gray radiation scheme (GR) and the comprehensive radiation scheme (RRTMG). The GR scheme is based on Frierson et al. (2006) and is very simple, with optical depths calculated as a function of pressure and latitude. The RRTMG scheme, based on the correlated-k method (Iacono et al. (2008)), is a more sophisticated radiation scheme with realistic representations of gases, clouds, and aerosols. The parameters of the GR scheme are tuned to match the global-mean SST and equator-to-pole SST contrast in the RRTMG scheme. The simulations were spun up for 10 years in the GR scheme and 20 years in the RRTMG scheme. Ten years

of 4x daily output were saved for both radiation schemes, provided courtesy of Z. Tan. Additional details of the simulations can be found in Tan et al. (2019).

2.1.3 ECMWF ERA-Interim Reanalysis

Observational analyses were performed using the ERA-Interim Reanalysis product from the European Center for Medium Range Weather Forecasts (ECMWF). Data is available from January 1, 1979 until August 31, 2019 in a variety of temporal and spatial scales. For this work we used both monthly-mean and daily-mean SAT values from 1979 to 2017, obtained from the ECMWF website (https://apps.ecmwf.int/datasets/data/interim-full-daily/levtype=sfc/). ERA-Interim reanalyses are produced using available observational data combined with a forecast model to estimate the state of the atmosphere Dee et al. (2011). The model also estimates various physical parameters. While the output is not fully observational, the parameters are constrained by the observations used for the forecast. Additional details can be found in Dee et al. (2011).

2.2 DYNAMICAL ADJUSTMENT METHODS

Dynamical adjustment seeks to determine the role of temperature variability that is solely due to internal atmospheric circulation changes, all other factors being equal. The overall model for dynamical adjustment is as follows:

$$T = T_{dyn} + \epsilon \tag{2.1}$$

where T is the temperature time series for any grid point or spatial region, T_{dyn} is the temperature predicted by internal circulation dynamics, and ϵ is the error (or the non-dynamical contribution to temperature variability, i.e., all other contributions to internal variability as well as the radiative signal). Chapter 3 examines various dynamical adjustment methods used to estimate T_{dyn}, the details of which are described below. Specifically, our analyses focus on the Northern Hemisphere extratropics (20N to 90N) during the boreal winter months (December through February, or DJF), as the amplitude of the atmospheric circulation is strongest during this time period. All analyses are performed on monthly mean variables.

2.2.1 Principal Component Regression (PCR)

The principal component method used for dynamical adjustment in Chapter 3 calculates T_{dyn} by selecting the first n principal component time series (PCs) of SLP and regressing them onto the temperature record. Principal Component analysis, also known as Empirical Orthogonal Function (EOF) analysis, decomposes a 2-D matrix into a set of orthogonal basis functions. EOF analysis is widely used in climate science to identify and investigate the primary modes of variability (e.g., the Arctic Oscillation Thompson and Wallace (1998); the Pacific Decadal Oscillation, Mantua et al. (1997)).

Given a matrix $\mathbf{X}$ with dimensions [M x N], the EOF and PCs are found by computing the eigenvectors and eigenvalues on a covariance matrix $\mathbf{C}$, where:

$$\mathbf{C} = \overline{\mathbf{X}\mathbf{X}^T} \tag{2.2}$$

and

$$\mathbf{C}\mathbf{E} = \mathbf{E}\Lambda \tag{2.3}$$

Solutions yield eigenvectors $\mathbf{E}$ (the EOFs) and their corresponding eigenvalues λ (diagonal values of Λ that represent the variance explained by the eigenvectors). The principal components can be found by projecting the original matrix $\mathbf{X}$ onto the eigenvectors:

$$\mathbf{Z} = \mathbf{X}^T\mathbf{E} \tag{2.4}$$

The matrix $\mathbf{Z}$ contains N principal components of length M from the original dataset. Like the EOFs, the PCs are mutually orthogonal. For our data, M is the time dimension, and N is the spatial dimension. Data are first spatially weighted by the cosine of latitude in order to account for the spherical shape of the earth.

To use PCR for dynamical adjustment on temperature, we write the method as follows:

$$T_{dyn} = \sum_{i=1}^{n} \alpha_i z_i \tag{2.5}$$

where the subscript i represents the corresponding EOF/PC mode and z represents the PC time series of SLP. If we assume that the principal components of SLP and temperature are linearly related, α is calculated simply as the regression coefficient:

$$\alpha_i = \frac{\overline{z_i' T'}}{\overline{z_i'^2}} \tag{2.6}$$

In Chapter 3, we apply the PCR method to the SLP field. The EOFs thus show spatial patterns that explain the highest variability in SLP, while the principal components are the corresponding timeseries. Several previous studies have applied PCR to calculate the dynamically-adjusted temperature field (e.g., Thompson et al. (2009); Smoliak et al. (2015); Saffioti et al. (2016)) using different numbers of principal components as predictors. In this work, we choose to use the first 10 PCs from the SLP field to represent the atmospheric circulation.

2.2.2 Maximum Covariance Analysis (MCA)

Maximum covariance analysis, as its name suggests, identifies linearly related patterns that has the greatest covariance between two datatsets (von Storch and Zwiers (1999)). While EOF analysis is used to identify patterns of variability in a single field, MCA is used to identify coupled patterns of variability between two fields. The calculation is as follows:

First, just as in EOF analysis, calculate the covariance matrix of $\mathbf{X}$ and $\mathbf{Y}$ (both M x N matrices), where M represents the temporal dimension and N the spatial dimension:

$$\mathbf{C}_{XY} = \overline{\mathbf{X}\mathbf{Y}^T} \tag{2.7}$$

This method can also be applied to maximize correlation instead of covariance between the two matrices. In this case, $\mathbf{X}$ and $\mathbf{Y}$ are first standardized before computing Equation 2.7, and $\mathbf{C}_{XY}$ is a correlation matrix.

Let **u** be an arbitrary unit column vector representing a pattern in **X**, and **v** be an arbitrary unit column vector representing a pattern in **Y**. Thus, the projections onto their respective data fields are:

$$\mathbf{a} = \mathbf{u}^T \mathbf{X} \tag{2.8a}$$

$$\mathbf{b} = \mathbf{v}^T \mathbf{Y} \tag{2.8b}$$

The MCA method decomposes the cross covariance matrix to find optimal patterns **u** and **v** that maximize their covariance:

$$\begin{aligned} c &= \mathrm{cov}[\mathbf{a}, \mathbf{b}] \\ &= \mathbf{u}^T \mathbf{C}_{XY} \mathbf{v} \end{aligned} \tag{2.9}$$

The maximum c can be found from the SVD of $\mathbf{C}_{XY}$:

$$\mathbf{C}_{XY} = \mathbf{U} \Sigma \mathbf{V}^T \tag{2.10}$$

The MCA method has been promoted in climate science research by Bretherton et al. (1992) and Wallace et al. (1992), who used MCA to analyze coupled patterns of 500mb height and SST variability. In Chapter 3, we apply the MCA method in two ways: 1) using the SLP field to find the spatial pattern that predicts the largest variance in the temperature field, and 2) using only the temperature field, where we find the spatial temperature pattern that has the greatest covariance with the temperature timeseries at every grid point.

2.2.3 Partial Least Squares (PLS)

Partial least squares (PLS) is a matrix method that uses a set of independent variables (**X**) to predict one or more dependent variables (**Y**). This relationship is predetermined; this is in contrast to the MCA method, which makes no distinction between predictor and predictand. The PLS predictors are generally ordered by the amount of variance they explain in the dependent variable.

To calculate the predictors, we find two sets of weights (**u** and **v**) that, when projected onto **X** and **Y**, maximize their covariance.

$$\mathbf{a} = \mathbf{X}\mathbf{u} \tag{2.11a}$$

$$\mathbf{b} = \mathbf{Y}\mathbf{v} \tag{2.11b}$$

such that

$$\mathbf{a}^T\mathbf{a} = 1 \tag{2.12a}$$

$$\mathbf{u}^T\mathbf{u} = 1 \tag{2.12b}$$

$$\mathbf{a}^T\mathbf{b} = \max \tag{2.12c}$$

Thus, the first part of PLS regression is the same as in the MCA method. After the weights and predictors are found, the PLS predictor (a_i) is subtracted from both **X** and **Y** using ordinary least squares regression. Then, the procedure is repeated with the new **X** and **Y** matrices to obtain the second PLS predictor.

PLS regression has been used in Smoliak et al. (2010) to examine 1) North Pacific SLP patterns associated with the variability in winter snowpack in the Cascade mountains; and 2) tropical SST patterns that modulate the year-to-year variations in Atlantic hurricanes. More recently, PLS regression has been used to investigate the influence of SLP on surface temperatures in observations (Wallace et al. (2012); Smoliak et al. (2015)).

In Chapter 3, PLS regression is used on grid point SLP and wind at 950hPa to predict surface temperature. The temperature time series is correlated with the standardized predictor (either SLP or wind) to form a cross-correlation pattern. Next, the standardized predictor field, weighted by the cosine of latitude to account for area weighting, is projected onto this pattern to obtain a predictor time series. This is then regressed onto the original temperature time series to obtain the first mode of T_{dyn}. This predictor is then subtracted out from both the original predictor field X and the predictand field Y. The analysis then starts over again with the new datasets. Each pass of

PLS regression thus removes more variability from the original datasets; a limited number of predictors (1-3 in this study) are retained to prevent overfitting. The final dynamical contribution to temperature is the sum of the T_{dyn} results from the assigned number of predictors. In the case when we apply PLS regression to both the zonal and meridional winds at 950hPa, the final T_{dyn} is calculated via multiple regression. Further details of PLS regression can be found in Smoliak et al. (2015).

2.2.4 Constructed Analogues

This method makes use of "analogs" in the predictor field to estimate patterns of variability in the predictand field. Circulation analogs have been used in weather prediction (Lorenz (1969); Van den Dool (1994); Van den Dool et al. (2003), and more recently has been applied to examine the contribution of dynamics to various temperature trends (Deser et al. (2016); Lehner et al. (2017); Merrifield et al. (2017)).

In Chapter 3, we apply the constructed analog method on the predictor field SLP. For both the control and the large ensemble runs, we obtain analogs from the control run, ranked by closest Euclidean distance. The advantages for using analogs from the control run are twofold: 1) the control run is not subject to any external forcings, and as such the analogs and the temperature patterns are solely from internal variability; and 2) the number of samples from which to draw analog is very large. We choose n number of the closest analogs ($\mathbf{X}_C$) and linearly combine them to form a constructed analog $\mathbf{X}_{CA}$. Given an initial dataset $\mathbf{X}_0$, a constructed analog is defined as:

$$\mathbf{X}_0 \approx \mathbf{X}_{CA} = \mathbf{X}_C \boldsymbol{\beta} \tag{2.13}$$

This results in a basic linear equation $\mathbf{A}x = \mathbf{B}$. However, because the elements of X are not orthogonal states no unique solution exists for the values of $\boldsymbol{\beta}$. We estimate the $\boldsymbol{\beta}$ coefficients by using the Moore-Penrose pseudoinverse, or by using SVD:

$$\boldsymbol{\beta} = [(\boldsymbol{X}_C^T \boldsymbol{X}_C)^{-1} \boldsymbol{X}_C^T] \boldsymbol{X}_0 \tag{2.14}$$

The $\boldsymbol{\beta}$ coefficients are then applied to the predictand field corresponding to the selected $\mathbf{X}_C$ analogs.

As an example, consider the case of calculating the dynamically-induced temperature for January 1955 in one CESM ensemble member ($\mathbf{X}_0$). First, we rank all January SLP fields in the control run (1800 samples) by closest Euclidean distance (i.e., the fields most similar to our target SLP field) and take the top 150 SLP analogs. We then randomly select 100 of these analogs to form $\mathbf{X}_C$. We find $\boldsymbol{\beta}$ using Equation 2.14 above, and apply the coefficients to the surface temperature anomaly data ($\mathbf{X}_{C,SAT}$) corresponding to the 100 SLP analogs selected (i.e., if the 100 SLP analogs came from years 201-300 in the control run, then we take temperature data from years 201-300) to calculate the dynamically-induced temperature component.

$$\mathbf{T}_{dyn} = \boldsymbol{X}_{C,SAT}\boldsymbol{\beta} \tag{2.15}$$

T_{dyn} is then the internal dynamical contribution on the January 1955 temperature field. For a more detailed description of applying SLP analogues to surface temperature, see Deser et al. (2016).

2.3 AUTOCORRELATION ANALYSES

Surface temperature can be modeled as a first order autoregressive (AR(1)) time series:

$$x(n_t) = b\,n_t + \alpha x(n_t - 1) + \epsilon(n_t) \tag{2.16}$$

where $n_t = 1, 2, ..., N$ is the number of time steps, b is the linear trend on the time series, $x(0) = 0$ by assumption, and ϵ is white noise (independent Gaussian noise with mean of zero and variance of σ_ϵ^2). The parameter α is between -1 and 1 and represents the memory in the time series $x(n_t)$ from one time step to the next. Higher α values represent larger memory in the timeseries from one time step to the next (e.g., the temperature at a given day is strongly correlated to the temperature from the day before).

We calculate the autocorrelation parameter α as follows:

$$\alpha_n = \frac{\overline{T(x)'T(x-n)'}}{\sigma_{T(x)}\sigma_{T(x-n)}} \tag{2.17}$$

where n is the lag time. This is identical to calculating the correlation coefficient, where the two variables are temperature and temperature lagged by n.

2.3.1 Fisher Z-transformation

In Chapter 4, we calculate the differences and the ensemble and zonal means of the autocorrelations from individual ensemble members. Because correlations range from -1 to 1, to perform any calculations we must first transform the distribution of our autocorrelations so that it becomes normally distributed. One way to do so is by using the Fisher Z-transformation, which is as follows:

$$z = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right) \tag{2.18}$$

The z in Fisher Z-transformation stands for the z-score. By transforming the sampling distribution to a Gaussian distribution, we can now calculate differences, averages, and confidence intervals on correlations. We also use this to calculate if two correlations are statistically significantly different from one another.

2.3.2 Spectral Analyses

Fourier transformation is used in Chapter 4 to filter the temperature time series. Filtering is often used to remove higher or lower frequencies, or to identify frequencies associated with certain signals. Fourier transform decomposes a function into a series of sines and cosines:

$$y(t) = A_0 + \sum_{k=1}^{N/2} A_k \cos(2\pi k \frac{t}{T}) + \sum_{k=1}^{N/2} B_k \sin(2\pi k \frac{t}{T}) \tag{2.19}$$

where $y(t)$ is a function of time from 0 to T, the length of the time series is N + 1, and A_k and B_k are the regression coefficients for the predictors.

The variance explained by a particular harmonic k is:

$$R^2 = \frac{A_k^2 + B_k^2}{2} = \frac{C_k^2}{2} \tag{2.20}$$

given that C_k^2 is defined as $A_k^2 + B_k^2$. The power spectrum is generally plotted as $C_k^2/2$ versus k.

In Chapter 4, we show the power spectra on daily surface temperature anomalies from the CESM-LE model as a function of the period in years. Results were calculated using Fourier transform on a Hamming window. A window function is applied to the data to avoid spectral leakage due to the discrete nature of the input data. The Hamming window is defined as:

$$w(n) = a_0 - (1 - a_0) * \cos\left(\frac{2\pi n}{N}\right) \tag{2.21}$$

where $a_0 = 0.54$, and $0 \leq n \leq N$. This window has a maximum at the middle and symmetrically tapers away at the ends. The temperature time series anomalies for each individual ensemble member is multiplied by Equation 2.21, and a Fourier transform is applied to the result to obtain the power spectrum.

3 Assessing Methods for Isolating the Signature of Internal Variability in Climate Change[1]

Internal variability plays a large role in determining surface climate on local and regional scales; as such, understanding the role of internal variability is crucial for accurately assessing and attributing climate trends. In this study we explore how to best remove dynamical aspects of internal variability in order to better pinpoint and understand the radiatively forced signal. Dynamical adjustment seeks to separate out the internal dynamical contribution to temperature trends, thus reducing the amplitude of internal variability that obscures the signal of anthropogenic forcing. We present a comprehensive analysis and comparison of multiple methods that have been used in the climate literature. Methods are first compared using a preindustrial control run, then assessed under a large ensemble of climate change simulations. The use of a control run tests the accuracy of the forced signal after dynamical adjustment, while the large ensemble explores the utility of each method in the presence of climate change. We present an alternative method of using partial-least-squares from the wind field and note its advantages to other comparable methods. These novel quantitative results help elucidate the role and details of each method and provide readers with method recommendations for future use.

3.1 INTRODUCTION

Given the large uncertainty in climate projections due to internal variability, there has been a long history of attempts to isolate the internal variability in both observations and climate change simulations. One major approach has been dynamical adjustment, which seeks to remove the variability in temperature due to internal circulation dynamics. While the circulation itself may respond to anthropogenic forcing, this response is generally much weaker than the surface temperature response (Deser et al. (2012b)).

[1]This chapter contains material that will be submitted.

Past studies have used a variety of methods to examine the role of internal circulation dynamics on surface climate over different regions and time periods, attempting to explicitly isolate the contribution to surface temperature from circulation dynamics. All these methods are variants of linear regression, where temperature is fitted to the dynamical component of internal variability, following the general equation below:

$$\mathbf{Y} = \alpha \mathbf{X} \tag{3.1}$$

where $\mathbf{X}$ is the time series representing internal circulation dynamics, $\mathbf{Y}$ is the contribution of $\mathbf{X}$ to any temperature time series, and α is the regression coefficient. The differences between most methods lie in how the predictor $\mathbf{X}$ is estimated. Generally, a spatial pattern over an area of interest is defined, and the change or index of that pattern over time is fitted to the temperature time series in that region. Wallace et al. (1995) found that approximately half of the variance of monthly mean temperature in the Northern Hemisphere was linearly related to a distinctive spatial pattern they named COWL (cold-ocean-warm-land). As its name suggests, COWL refers to the dynamical effect that when the oceans are anomalously cold, the land regions are anomalously warm. Other early studies used indices of well known large-scale dynamical patterns. Hurrell (1996) used multivariate linear regression to show that much of the temperature trends in the Northern Hemisphere from the mid 1970s to early 1990s were due to natural variability patterns (rather than anthropogenic forcing). For instance, changes in the North Atlantic Oscillation (NAO) resulted in cooling in the northwest Atlantic and warming across Europe and Eurasia, while teleconnections from ENSO and its feedbacks in the midlatitudes were linked to temperature changes in the Pacific basin and North America. Quadrelli and Wallace (2004) applied principal component based indices from the sea level pressure field (the NAM and the PNA patterns) to surface air temperature trends over the time period 1958-1999. Thompson et al. (2009) used maximum covariance analysis (MCA), where $\mathbf{X}$ is calculated as the spatial SLP pattern that explains the greatest variance in global mean temperature. A similar method, partial least squares (PLS), uses correlation instead of covariance to calculate anomaly patterns that are tailored to maximally explain concurrent variations in surface temperature (e.g., Wallace et al. 2012; Smoliak et al. 2015). Deser et al. (2016)

used a variation of the constructed analog method developed in Van den Dool et al. (2003) to estimate dynamical adjustment in temperature trends over North America for the past 50 years. This constructed analog method looks for similar SLP patterns in a preindustrial control run, and using a linear combination of these patterns to calculate a constructed SLP analog **X**. Regression coefficients for each analog pattern are then calculated and used to estimate the dynamically-induced temperature time series **Y**.

The multiple options available for dynamical adjustment naturally leads to the question “which method is best?" This is not an easy question to answer, given that the definition of “best" changes in different situations. While many previous studies have examined the effects of internal atmospheric dynamics on temperature trends in observations and models, no study has specifically focused on an in-depth comparison between these different methods and their respective biases. Furthermore, most previous studies have examined dynamical adjustment methods in a single model or in the observational record, which only provide a single realization of reality and is not able to adequately estimate the forced signal nor quantify its internal variability. This chapter seeks to provide a comprehensive and quantitative analysis and comparison of different methods present in the climate literature, providing a helpful reference for dynamical adjustment methods and the contexts in which they might be used.

3.2 RESULTS

3.2.1 Comparing methods using a preindustrial control run

We define two main objectives for all dynamical adjustment methods: 1) to remove as much internal variability as possible, and 2) to remove as little of the forced signal as possible. These will thereafter be referred to as Objective 1 and Objective 2, respectively. For this study, we define the “best" method as the one that comes closest to meeting both objectives (i.e., the method that brings the temperature trends closest to the true radiative forcing while removing the most internal variability). In this study, we focus exclusively on the winter season (December-February, DJF) over the Northern Hemisphere extratropics (north of 20° latitude).

We first assess the methods using a 1800-year preindustrial control from the NCAR CESM1 model. As the control run has constant external forcings (no anthropogenic forcings or natural external forcings such as volcanic eruptions or solar irradiance changes), the only variability present within this run is internal variability (see Figure 2.1 for a sample of the surface temperature timeseries). Since there is no forced signal, there are also no feedbacks due to the forced signal (i.e., the magnitude of internal variability stays constant). This known true forced signal of zero allows for a quantitative assessment of Objective 2. The long simulation period also lends confidence to the robustness of our results. Thus, the control run allows for an explicit comparison between the different methods of estimating dynamical adjustment.

To observe our control run results in the context of the observational record and the large ensemble, we randomly select 10 different non-overlapping 50-year periods from the control run. Figure 3.1 shows the 50-year wintertime SAT and SLP trends for each individual "ensemble member." Monthly-mean data are averaged to form a seasonal timeseries, and the trend is then calculated via linear regression. The bottom right panel (labeled EM) shows the average of these 10 different 50-year trends, which are very close to zero. This is expected, since over a long time period the internal variability will average out, leaving only the radiative signal. However, even with no external forcings, the 10 random 50-year periods from the control run show significant trends in both surface temperature and SLP. Because we know that these trends are entirely due to internal variability, dynamical adjustment allows us to identify the role of circulation dynamics on surface temperature. This is a clear advantage of the control run and is not possible in a large ensemble with external forcings; even though the differences between the ensemble members in the climate change simulations are solely due to internal variability, we must also account for internal variability changes with climate change (if any), dynamical changes with climate change (if any), and changes in the relationship between internal circulation dynamics and surface temperature with climate change (if any).

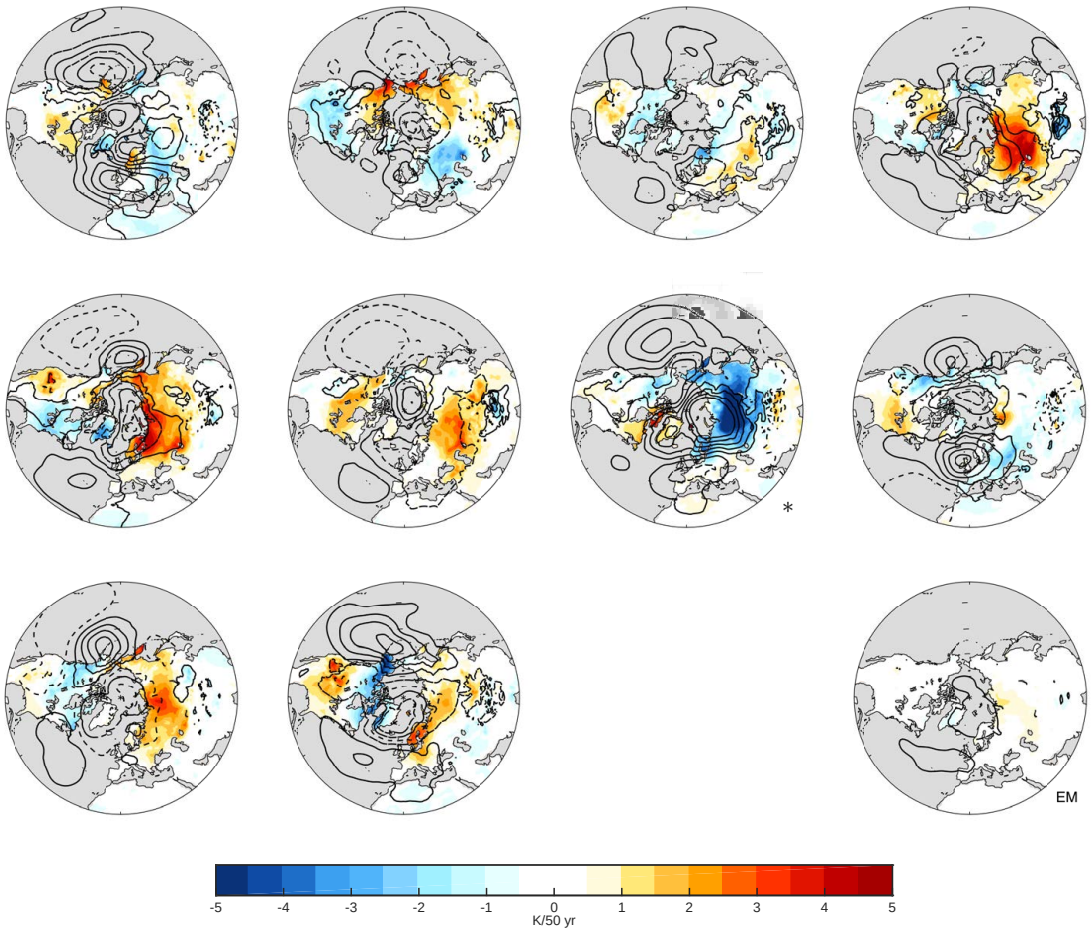

Fig. 3.1. 50-year temperature (color shading; K/50 yr) and SLP (contours; hPa/50 yr) trends from 10 different "ensembles" randomly selected from the 1800-yr control run. The SLP contour interval is 1.5hPa, with solid (dashed) contours for positive (negative) values. Temperature trends range from -5 to 5 K/50 yr. The bottom right panel (labeled EM) shows the average temperature and SLP trends of the 10 different 50-year periods shown.

To provide a quantitative comparison between methods, we use the metric root-mean-square-error (RMSE). RMSE is defined as the square root of the second sample moment of the differences between the predicted values and observed values and can be written as:

$$RMSE(x) = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \tag{3.2}$$

where $\hat{y}_i$ is the observed temperature trend at a single gridpoint i, y_i is the true temperature trend (zero in the control run), and n is the sample size (here, all grid points between 20N and 90N).

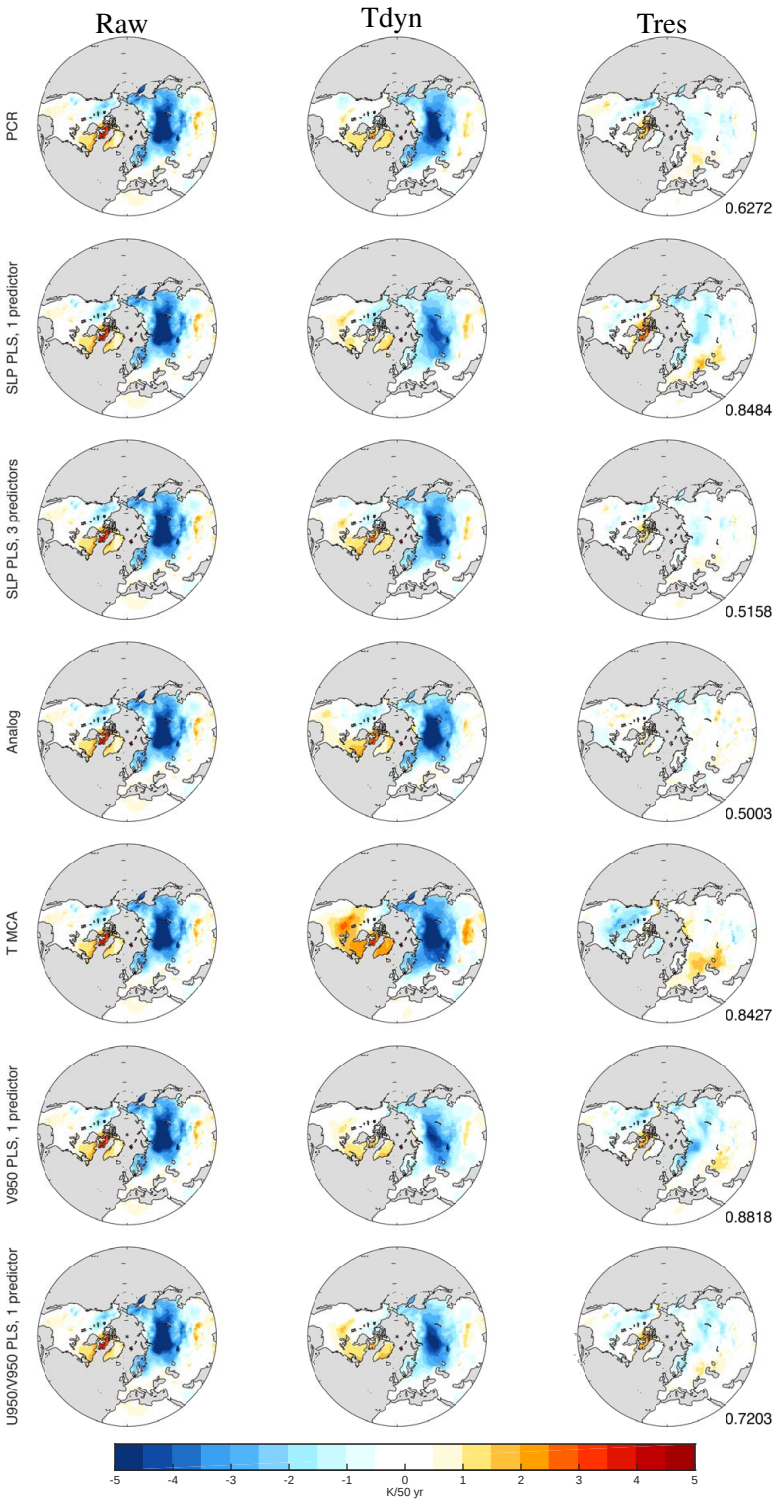

Fig. 3.2. 50-year DJF temperature trends from the starred 50-year period in Figure 1 for (left) raw data from the control run, (middle) the dynamical contribution on temperature, and (right) the dynamically-adjusted temperature from 7 different dynamical adjustment methods. The numbers on the dynamically-adjusted temperature trends show the RMSE values for that method. See text for details.

Thus, the RMSE equation results in a single value per ensemble member. A lower RMSE value represents a smaller deviation from the true forced signal, and as such is able to provide a quantitative comparison metric for Objective 2. However, the RMSE does not provide any information on Objective 1, as a method may estimate the forced signal well but not remove any variability. If the observed and predicted values match perfectly, the RMSE results would be zero, which serves as the baseline for our comparisons. We apply this RMSE metric to the raw and residual ("Tres") temperature trends from each of the dynamical adjustment methods.

We note that almost all the previous methods in the literature are based on SLP. While SLP is a useful metric for estimating the circulation, here we propose using the near-surface wind field in lieu of SLP. In specific, we will also compare additional methods of PLS using the zonal and meridional winds at 950hPa. The wind field has the advantage of being directly in the temperature advection equation, and thus provides a more direct relationship to the temperature field. We introduce two additional methods: a PLS method that uses V950 (both 1 and 3 predictors to directly compare to SLP PLS results), and a PLS method that calculates 1 predictor for both U950 and V950, and then combined to form a single predictor using multiple regression. These new methods will be assessed along with all the methods previously listed in both the preindustrial control run and the CESM large ensemble. We will also discuss the response of the SLP and wind fields to climate change.

We apply the dynamical adjustment methods to all 10 ensembles from the control run, resulting in a timeseries of temperature due to internal dynamics ("Tdyn") and a residual temperature timeseries ("Tres"). Figure 3.2 shows the raw, dynamically-induced, and dynamically-adjusted 50-year temperature *trends* from applying each of the methods to a single 50-year period from the control, the starred ensemble member from Figure 3.1. From Figure 3.1 we see that the temperature trends in this 50-year period include a very strong cooling trend across most of Russia, eastern Europe, and Scandinavia (the trends are reproduced in the left column of Figure 3.2). The middle column of Figure 3.2 ("Tdyn") shows the 50-year dynamically-induced temperature trends calculated using each method, while the right column ("Tres") shows the dynamically adjusted trends. Comparing the left and middle columns, we see that almost all of this cooling trend is due to internal

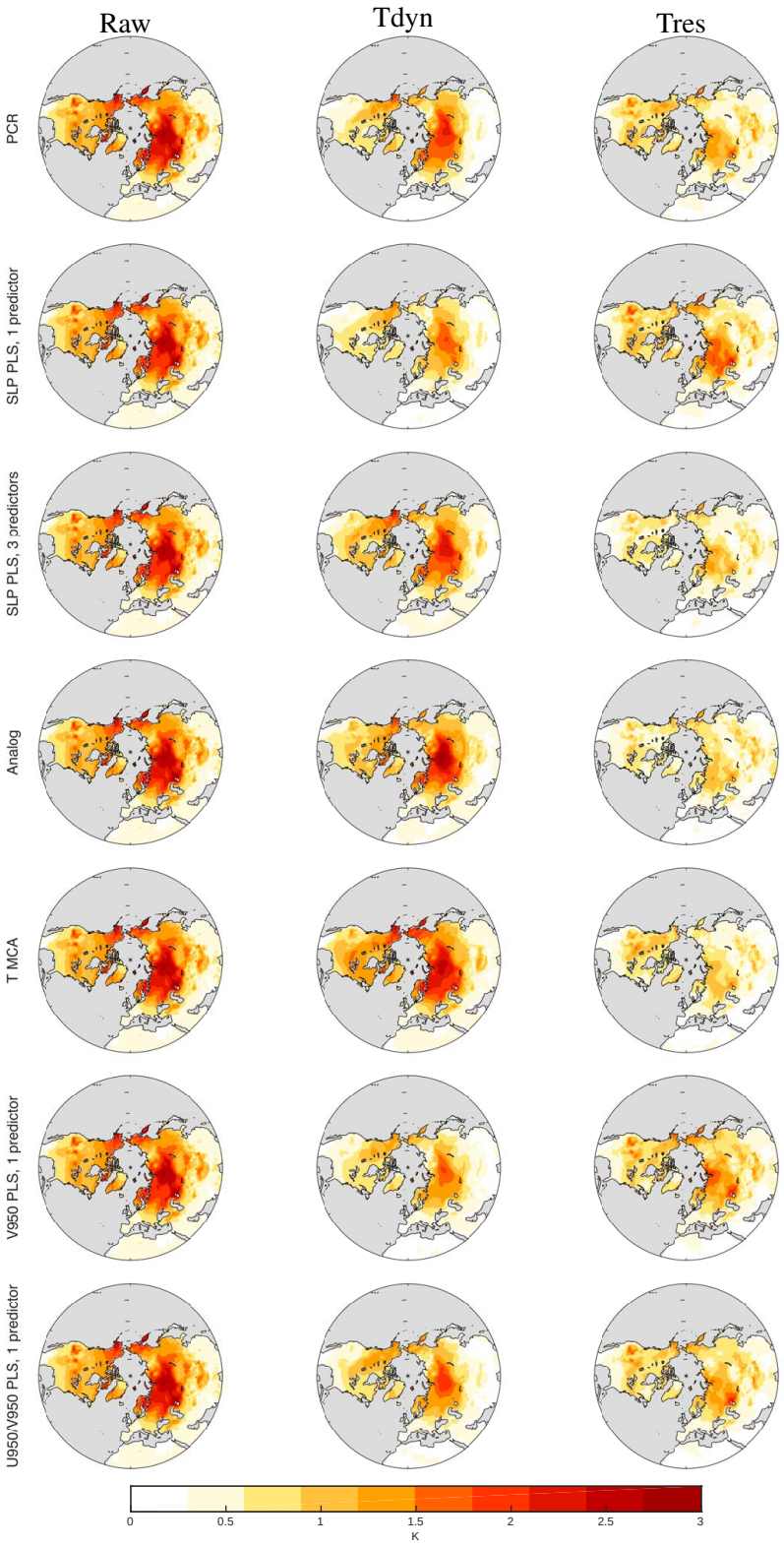

Fig. 3.3. Standard deviation on 50-year DJF temperature trends across all 10 ensembles from the control run for (left) raw data, (middle) dynamically-induced temperatures, and (right) dynamically-adjusted temperatures.

atmospheric dynamics. Moreover, the methods are all very consistent in picking up this cooling, along with a smaller warming trend in Greenland and northeastern Canada. Despite the consistencies, slight regional differences amongst the methods can be seen from the middle and right columns. Results from the temperature-based method show a stark difference between this method and ones based on the SLP field. The temperature-based method (T MCA) attributes a larger cooling trend in eastern Europe and Russia to internal dynamics, as well as a stronger warming trend over North America that the circulation-based methods lack. The numbers in the right panel of Figure 3.2 show the respective RMSE values of the dynamically-adjusted trends. For reference, the raw RMSE (the difference between the left column of Figure 2 and zero, without any dynamical adjustment) is 1.9407. The RMSE values in all dynamical adjustment methods significantly decrease, thus bringing the residual trends much closer to the true trend of zero. Quantitatively, the constructed analog and SLP PLS with 3 predictors have the lowest RMSE values, which match the visual results.

To evaluate Objective 1 (to remove as much of the internal variability as possible), we calculate the standard deviation across all 10 ensemble members for the raw, dynamically-induced, and dynamically-adjusted temperature trends (Figure 3.3) by extending the analysis from Figure 3.2 to all 10 ensemble members for all methods. The left column shows the standard deviation on the raw temperature trends (e.g., the range of the different 50-year trends in Figure 3.1). Note that all panels in the left column are identical. The middle column (“Tdyn") shows the standard deviation of the dynamically-induced temperature trends, while the right column shows the standard deviation on the residual trends. While Figure 3.2 showed the temperature trend results in a single 50-year period, Figure 3.2 shows the range on the trends between all 10 “ensemble members." The largest variability in the SAT trends are over the higher latitudes, with Siberia, Alaska and northeastern Russia, and Scandinavia as the regions with the largest standard deviations. In general, variability increases with latitude, although regions such as western and central Europe and central Asia show surprisingly low variability in their temperature trends. All methods remove a significant amount (over 50%) of the variance in the temperature trends, with most of the dynamically-induced

variability in the higher latitudes, especially over Siberia and North America. The dynamically-adjusted residual trends (right column) have significantly less variability in comparison, and the spatial patterns are less distinct. From Figure 3.3, it appears that the SLP PLS with 3 predictors, constructed analog, and T MCA methods remove the most variability.

Table 3.1. Summary of RMSE results between the true forced signal (zero in the control run) and dynamically-adjusted temperature trends from each of the 10 individual ensemble members. The columns show the mean of the RMSE results, the standard deviation, and the single highest RMSE across all 10 ensemble members, respectively. Results are shown for all dynamical adjustment methods applied to the 10 different 50-year periods from the control run.

Method	**Mean RMSE**	**Standard Deviation**	**Maximum**
None (unadjusted)	1.2099	0.3521	1.9407
PCR (10 PCs)	0.7358	0.0893	0.8437
SLP PLS (1 predictor)	0.8239	0.1397	1.0184
SLP PLS (3 predictors)	0.5965	0.0880	0.7252
Constructed Analog	0.6108	0.0773	0.6946
T MCA (corr)	0.6232	0.0933	0.8427
V950 PLS (1 predictor)	0.8316	0.0901	0.9567
V950 PLS (3 predictors)	0.5219	0.0671	0.6412
U950 + V950 PLS (1 predictor)	0.7618	0.1134	0.9033

Table 3.1 summarizes the RMSE results for all 10 ensemble members from the control run. Multiple ensemble members allow us to assess the consistency of RMSE results and thus provide a more robust conclusion. The mean, standard deviation, and maximum RMSE values of the 10 ensemble members are given for each dynamical adjustment method. The mean RMSE is the

Table 3.2. RMSE results between the true forced signal (zero in the control run) and the ensemble-averaged dynamically-adjusted temperature trends. Results are shown for all dynamical adjustment methods applied to the 10 different 50-year periods from the control run.

Method	**RMSE**
None (unadjusted)	0.3137
PCR (10 PCs)	0.1995
SLP PLS (1 predictor)	0.2008
SLP PLS (3 predictors)	0.1507
Constructed Analog	0.1753
T MCA (corr)	0.1737
V950 PLS (1 predictor)	0.2438
V950 PLS (3 predictors)	0.1387
U950 + V950 PLS (1 predictor)	0.2035

best overall metric for Objective 2, as it directly compares the dynamically-adjusted trends to the true forced signal. The standard deviation and maximum values indicate the consistency of the method in different instances of internal variability. Since the observational record represents only a single realization and thus could fall within the range of the entire ensemble, we also seek a low maximum RMSE value. Thus, both the mean and variability in the RMSE results are important in evaluating the methods. The first row shows the RMSE results for the control run ensembles without dynamical adjustment, which provides a reference point for comparison. Table 3.1 suggests that the SLP and V950 wind PLS methods with 3 predictors perform the best with respect to Objective 2. The constructed analog method also does very well, with a higher mean RMSE but low standard deviation and maximum, suggesting a narrower distribution.

RMSE values are also shown for the ensemble-mean trends (Table 3.2). The first RMSE value corresponds to the raw unadjusted data and represents the difference between the bottom right panel in Figure 3.1 and zero. For the other rows, $\hat{y}_i$ is calculated as the average of the ten dynamically adjusted ensemble trends from each method. Table 3.2 shows that all dynamical adjustment methods significantly decrease the RMSE in the ensemble-mean. The results mirror those of Table 3.1, with the constructed analog, T MCA, and PLS methods with 3 predictors having the lowest RMSE values. Tables 3.1 and 3.2, together with Figures 3.2 and 3.3, paint a comprehensive picture of the usefulness of each method in light of our two objectives. When using a single PLS predictor, V950 winds and SLP perform about the same, while adding U950 winds improves the RMSE and variance results slightly.

3.2.2 Applying dynamical adjustment methods to the large ensemble

We next extend the analyses to the large ensemble simulations, which provide a novel framework in which to test the dynamical adjustments method results from the control run. The key difference between the control run and the large ensemble simulations is the added radiative forcing in all ensemble members. In this study, we use the first 35 simulations run at NCAR due to slight discrepancies in surface temperature in the last 5 simulations. The forced signal is not known with precision, but is well estimated as the ensemble-mean trend. With 35 ensemble members, we

are able to quantify internal variability as the differences between the ensemble members, which allows us to test our methods in the presence of external forcing. We apply the same dynamical adjustment methods from Section 3.2.1 to the CESM-LE simulations for the period 1955-2004. We chose this time period because the CESM-LE is run with historical forcings until 2005, which minimizes uncertainty due to future emission scenarios.

We first form 3-month winter averages for SLP and temperature at each grid point in each ensemble member for the period 1955-2004, and then calculate trends over the entire 50 year period. These 50-year trends for temperature and SLP are shown in Figure 3.4 for all 35 ensemble members, with the ensemble mean trends in the bottom right panel (labeled EM). Recall that differences between ensemble members are solely due to internal variability. Comparing Figure 3.4 to Figure 3.1, we note the clear presence of radiative warming in the large ensemble runs. While internal variability still results in significant differences in temperature trends amongst the ensemble members, there are few areas with cooling trends across the 35 ensemble members here. Instead, the differences between the ensemble members are mostly varying degrees of warming in different regions.

Figure 3.5 shows the dynamical adjustment results from a single 50-year period, ensemble member #10. The left panel shows the difference between the raw temperature trends and the ensemble-mean trends (both shown in Figure 3.4). Compared to the ensemble mean, ensemble #10 exhibits much stronger warming over eastern Europe through central Siberia, and stronger cooling over most of Canada and Greenland. The middle panel shows the dynamically-induced trends, while the right panels show the difference between the residual trends and the ensemble-mean trends. This is analogous to the right column of Figure 3.2, as we want the residual temperature trends to be as close to zero as possible. Figure 3.5 suggests that both the strong warming across eastern Europe and Russia and the cooling across Canada and Greenland are due to internal dynamics, leaving a residual cooling trend in Sibera from other factors (thermodynamic processes, internal dynamics not accounted for in dynamical adjustment, etc.). Interestingly, Figure 3.5 again suggests that all the methods agree on the general patterns associated with internal dynamics. The

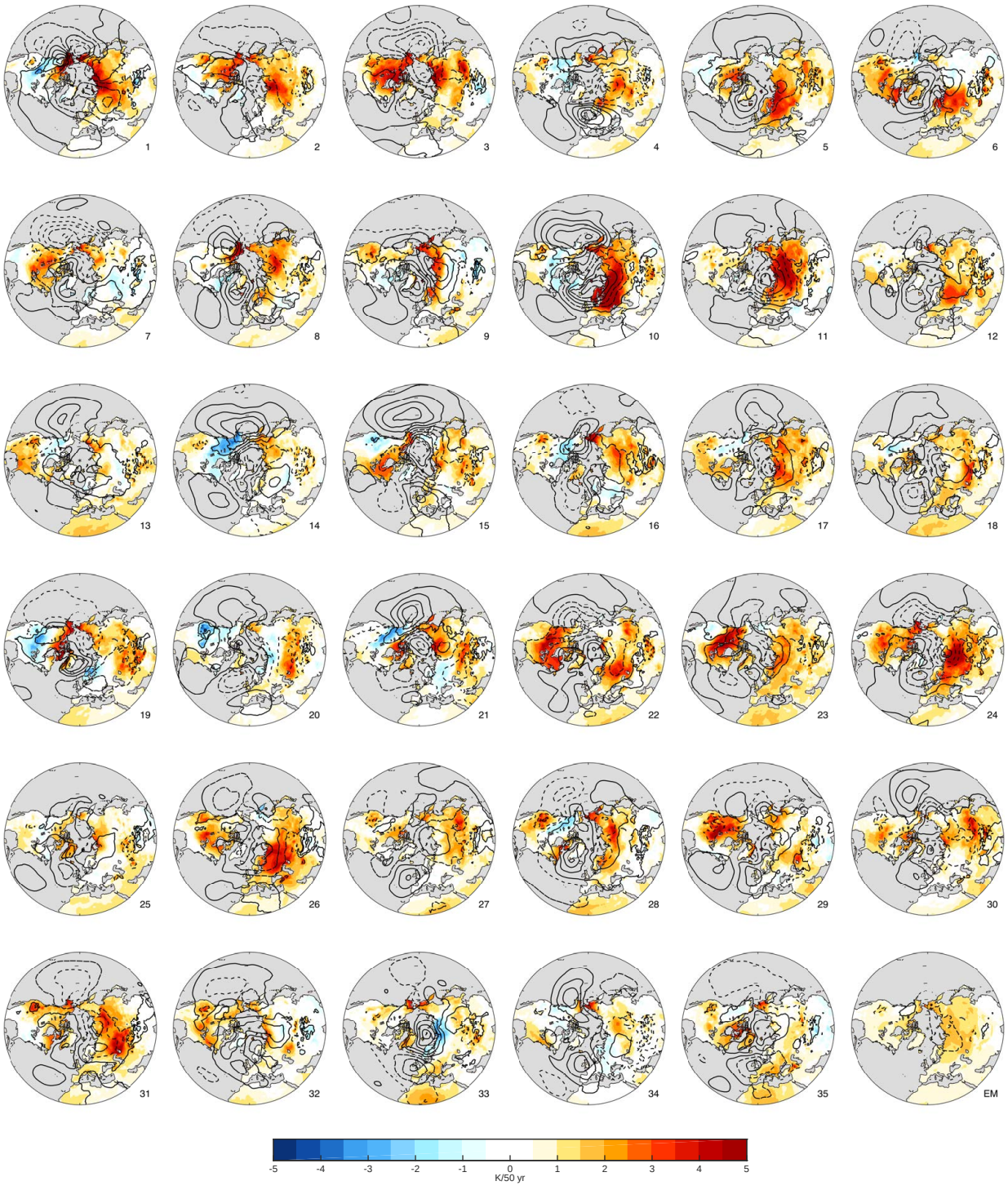

Fig. 3.4. 50-year DJF surface temperature (color shading; K/50 yr) and SLP (contours; hPa/50 yr) trends for each member of the CESM-LE (labeled 1-35) and the ensemble-mean (labeled EM) for the period 1955-2004. The SLP contour interval is 1.5hPa, with solid (dashed) contours for positive (negative) values.

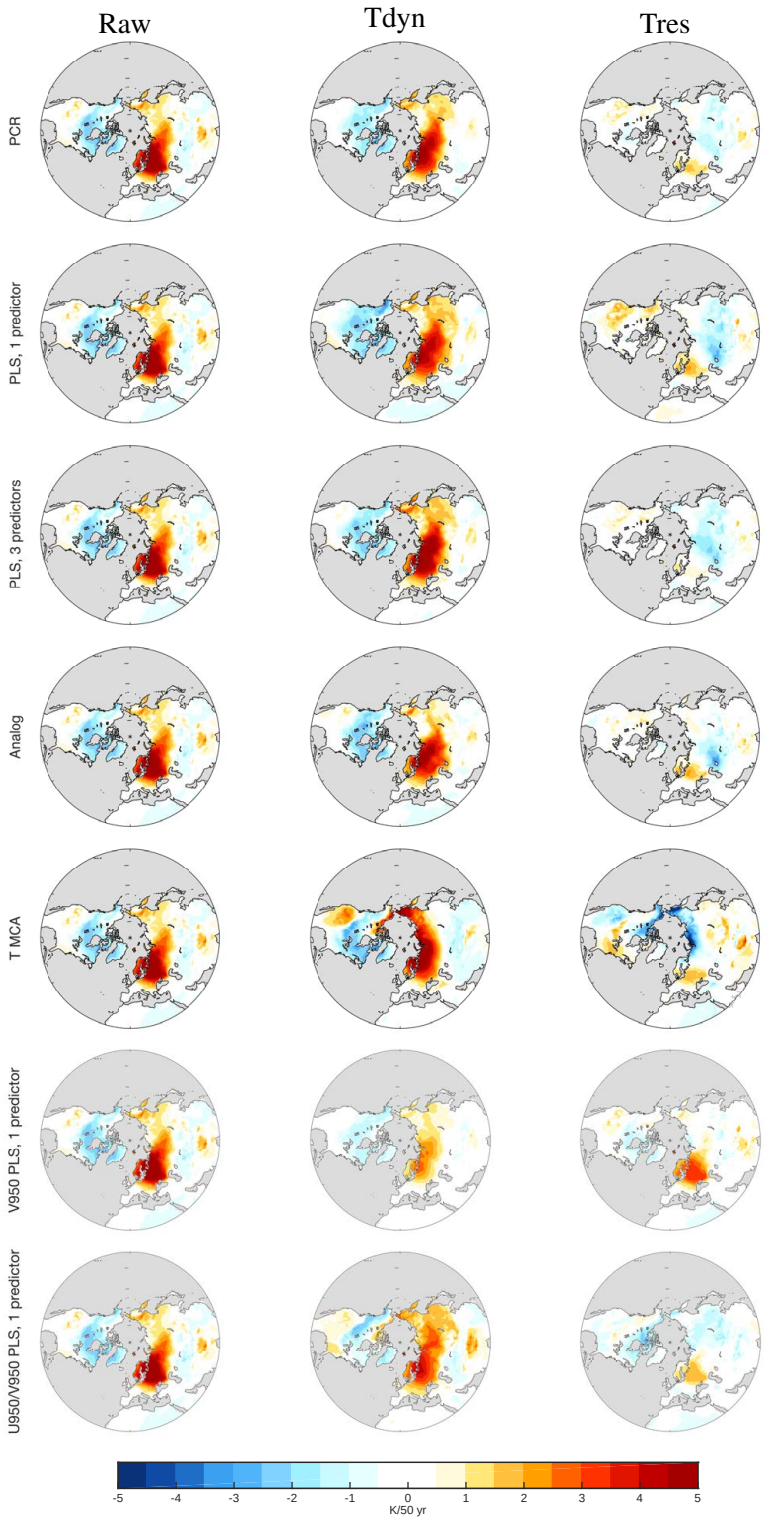

Fig. 3.5. Decomposition of 50-year wintertime temperature trends from ensemble member 10 of the CESM-LE. (left) difference between the raw temperature trends and the ensemble-averaged trends; (middle) dynamically-induced temperature trends calculated from the labeled method; (right) residual temperature trends from the ensemble-mean. Results are over the period 1955-2004.

strong warming trend over most of Russia and the cooling trend over Canada are clear in SLP, temperature and 950mb wind-based methods.

Figure 3.6 is analogous to Figure 3.3 and shows the standard deviations of dynamically-induced and dynamically-adjusted trends for each method, now in the large ensemble. We note a key difference: the left column, which represents the variability amongst the 35 ensemble members shown in Figure 3.6, is overall smaller than the left column in Figure 3.3 (variability amongst the 10 ensemble members in the control run). This is likely due to the larger sample size of the large ensemble; an F-test found no significant differences in internal variability between the control and forced runs (not shown). Dynamical adjustment consistently removes a large portion of the internal variability (approximately 50%), and the standard deviation on the residual trends (right column) are comparable to results from the control run.

Tables 3.3 and 3.4 are analogous to Tables 3.1 and 3.2 and summarize RMSE results from the large ensemble. The predicted values y_i are now the ensemble-averaged trends of the 35 unadjusted ensemble members, our best estimate of the true forced signal. Table 3.3 shows the distribution of individual ensemble RMSE results. Compared to the control run results, dynamical adjustment in the large ensemble has less of an effect in bringing trends closer to the forced signal. RMSE values for dynamically-adjusted trends in the large ensemble are larger and have a wider distribution. Most surprisingly, the wind (V950 and U950) based PLS methods give RMSE values that are much higher than expected from the control run results. Moreover, adding additional predictors on the first V950 PLS predictor does not improve RMSE results, but rather worsens them.

Because we assume the SLP field does not respond significantly to radiative forcing, the SLP-based methods are calculated in the same way as the control run. PLS with 3 predictors has the lowest RMSE values for all metrics, with the constructed analog method a close second, just as in the control run analyses. For the temperature-MCA method, we now must first remove the forced signal, lest it projects onto the dynamical adjustment calculations. We present three different options in the large ensemble: subtracting the ensemble-mean trend (labeled 'resid' in Table 3.3), detrending (a common metric for a single model output or observations, labeled 'detrend'), and removing the global-mean trend (here, the spatial average of our region of interest, between 20N

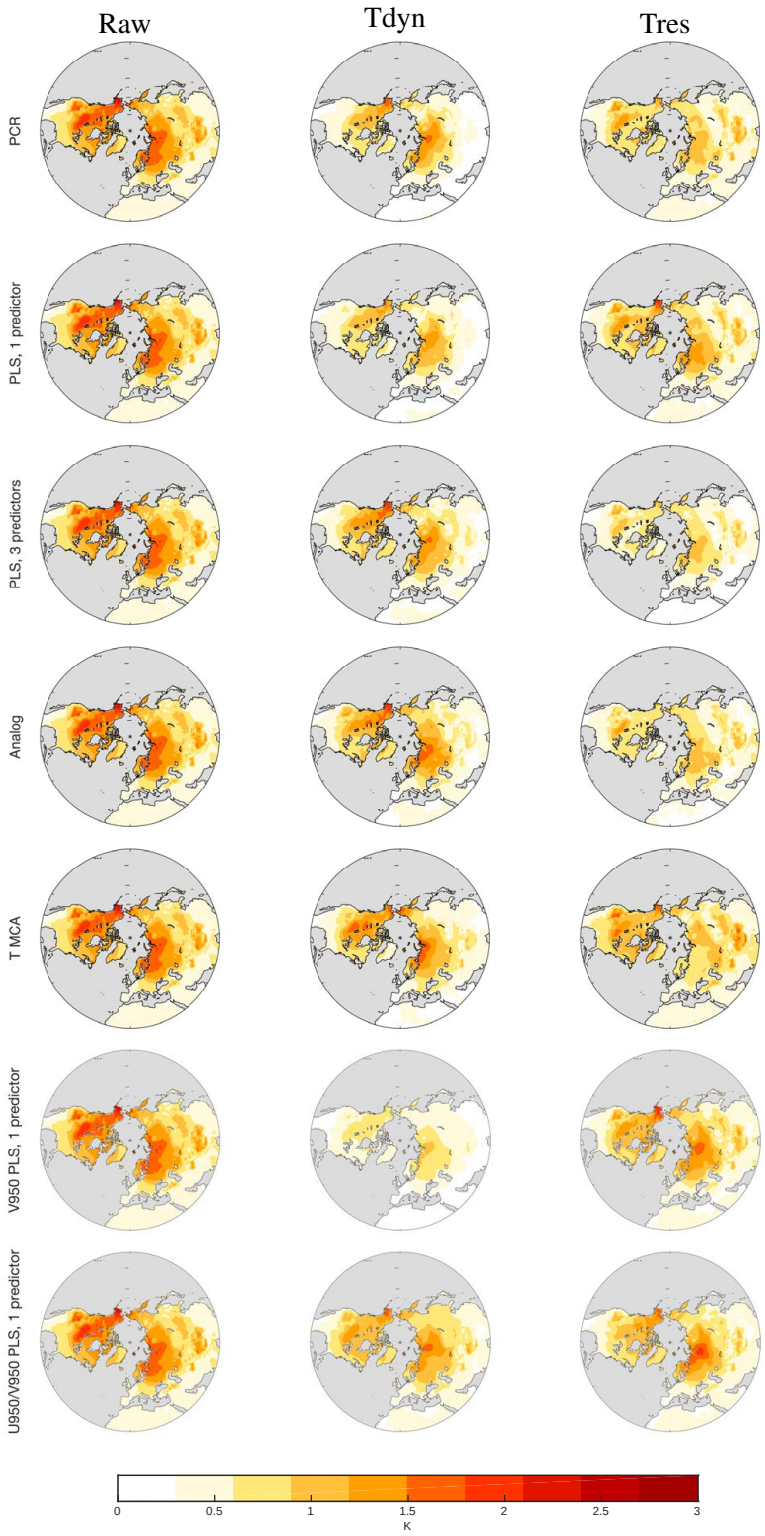

Fig. 3.6. As in Figure 3.3, but for the CESM-LE. Standard deviation on 50-year DJF temperature trends across 35 ensembles from the CESM-LE for (left) raw data, (middle) dynamically-induced temperature, and (right) dynamically-adjusted temperature for each dynamical adjustment method listed. Results are shown over the period 1955-2004.

and 90N, labeled 'global mean resid'). Results from subtracting the ensemble-mean trend are very similar to those in the control run; this is expected, since we have isolated the internal variability of the model. However, detrending or removing the global mean trend results in very high RMSE values. A simple linear trend or the global-mean trend does not estimate the forced signal well, and thus results in inaccurate estimates of the internal variability as well as projecting onto the dynamical adjustment.

RMSE results of the ensemble-mean also shows poorer results compared to the control run (Table 3.4). This is surprising, since the large ensemble has 35 members instead of 10, and a larger sample size is more likely to be more consistent and closer to the true forced signal. In specific, SLP PLS with 3 predictors went from the best method in the control to the worst method in the large ensemble. This result suggests that either 1) our estimate of the true forced signal is inaccurate, 2) the PLS method now produces a more inaccurate result due to external forcing projecting onto the SLP field, or 3) the relationship between SLP and internal circulation dynamics has changed and the methods cannot identify them as easily. The low RMSE value for the temperature-based MCA method lends support to the second hypothesis. In contrast, all SLP and wind-based methods have much higher RMSE values in the ensemble-mean, suggesting that external forcing affects these methods, leading to an inaccurate representation of both the dynamically-induced and dynamically-adjusted trends.

3.2.3 Examining biases in dynamical adjustment methods

To examine the discrepancies between the RMSE results, we look at possible biases in the dynamically-adjusted trends from projections of the forced signal onto all methods. As noted earlier, biases in methods can shed light onto the RMSE results, or provide additional information not shown in the RMSE calculations. Indeed, while the control run results showed no obvious biases, in the large ensemble all SLP-based methods seem to underestimate the forced signal - that is, temperature trends after dynamical adjustment are consistently smaller than the true forced signal. Figure 3.7 shows the difference between the forced signal (ensemble-mean trends of unadjusted temperature data) and the ensemble-mean T_{res} trends after dynamical adjustment for all methods.

Table 3.3. As in Table 3.1, but for the large ensemble. Summary of RMSE results between the estimated forced signal (raw ensemble-averaged trends) and dynamically-adjusted temperature trends from each of the 35 individual ensemble members in the large ensemble calculated using each of the above methods.

Method	Mean RMSE	Standard deviation	Maximum
None (unadjusted)	1.0066	0.2192	1.6169
SLP PCR, 10 PCs	0.7286	0.1143	1.0533
SLP MCA	0.8231	0.1498	1.1744
SLP PLS (1 predictor)	0.8058	0.1374	1.0433
SLP PLS (3 predictors)	0.6880	0.1043	0.9775
SLP Constructed Analog	0.7223	0.1043	0.9839
T MCA (corr, resid)	0.6208	0.0787	0.8644
T MCA (corr, detrend)	1.1403	0.1441	1.5717
T MCA (corr, global mean resid)	0.9775	0.1431	1.3058
V950 PLS (1 predictor)	0.8645	0.1364	1.1672
V950 PLS (3 predictors)	0.8435	0.1626	1.2007
U950 + V950 (1 predictor)	0.9523	0.2001	1.4143

Table 3.4. As in Table 3.2, but for the large ensemble. RMSE results between the estimated forced signal and the dynamically-adjusted ensemble-mean trends calculated using each of the listed methods.

Method	Mean RMSE
PCR (10 PCs)	0.2460
SLP MCA	0.2352
SLP PLS (1 predictor)	0.2738
SLP PLS (3 predictors)	0.3750
Constructed analog	0.3469
T MCA (reg)	0.0443
T MCA (corr, resid)	0.0421
V950 PLS (1 predictor)	0.2046
V950 PLS (3 predictors)	0.5567
U950 + V950 PLS (1 predictor, multiple regression)	1.0787

All SLP-based methods show a positive bias (the dynamically-adjusted trends underestimate the forced signal), likely due to the SLP ensemble-mean trend. The ensemble-mean SLP trend over 1955-2004 is overlaid onto the temperature trend differences. This bias is the likely reason for the high RMSE values in Table 3.4, a direct contrast to the results from the control run. This is especially interesting because PLS with one predictor shows a smaller bias than PLS with 3 predictors, suggesting that the addition of more predictors, while removing more internal variability, compromises on the accuracy of the estimate in the presence of external forcing. The constructed

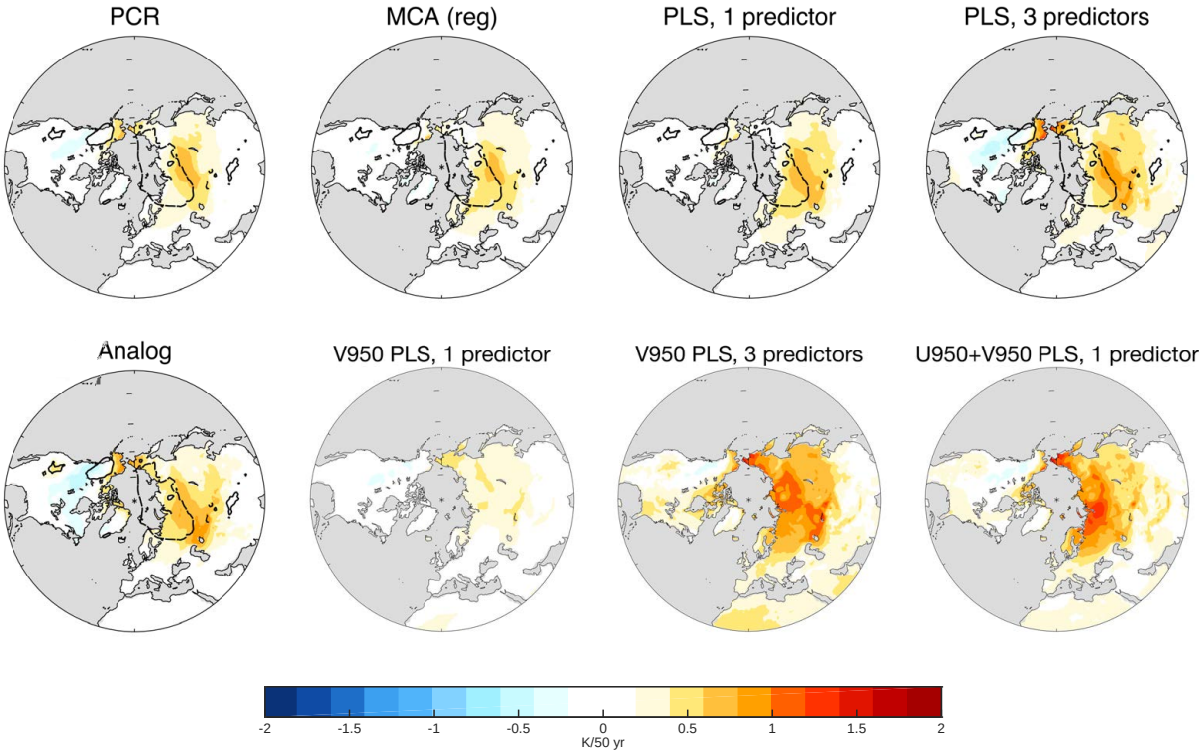

Fig. 3.7. Differences between the raw ensemble-mean temperature trends (the forced signal) and the dynamically-adjusted ensemble-mean temperature trends in the CESM-LE for 8 different dynamical adjustment methods. The ensemble-averaged 50-year SLP trend from the large ensemble is reproduced in methods based on SLP. See text for details.

circulation analog method specifically uses SLP data from the preindustrial control, and any change in SLP in the forced runs would have a tangible effect on this method's results. The wind-based PLS methods show very different results. V950 PLS with a single predictor shows almost no biases in its method, but when additional predictors are added (either additional V950 predictors or a U950 wind predictor combined using multiple regression) we see a significant difference in the ensemble-mean trends, mostly over Siberia. The temperature-MCA method (not shown) does not show any bias, suggesting that the forced signal did not project onto the temperature variability patterns themselves.

To test this possible cause of the biases, we examine the response of the SLP and near-surface wind fields to external forcing. Figure 3.8 shows the ensemble-averaged 50-year DJF trends in the CESM-LE for U950, V950, and SLP. The stippling shows grid points that are statistically significant from zero at the 95% confidence level. The wind fields both show very few areas where

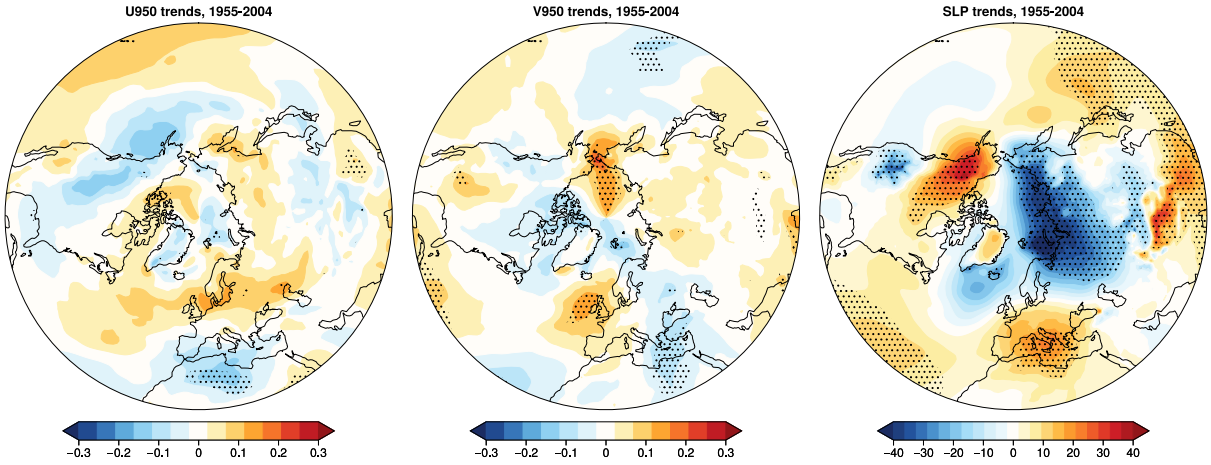

Fig. 3.8. 50-year DJF trends in the CESM-LE ensemble mean for (a) U950 winds, (b) V950 winds, and (c) SLP. Stippling shows grid points whose trends are statistically significant from zero at the 95% confidence level.

the trend is significant, but the SLP plot shows large areas of significant trends. The large regions across north and central Russia and northern Canada coincide with the bias regions in Figure 3.7. This SLP response is likely to increase as external forcing increases, suggesting that our assumption that SLP stays constant may not be valid. This could also be the reason for the discrepancy in the SLP-based methods between Tables 3.2 and 3.4. As a result of this SLP response to external forcing, the SLP-based methods are representing the warming from Figure 3.7 as part of internal dynamics, and are thus attributing them to the dynamically-induced temperature. Thus, the residual temperature trends are biased low and result in poorer RMSE values.

3.3 DISCUSSION

We also considered temperature-only methods, which theoretically would remove more internal variability than its SLP or wind-based counterparts due to the fact that temperature is better correlated with itself (by construction) We find that temperature MCA methods indeed yield a lower RMSE than their corresponding SLP methods in the control run. It also performs well in the large ensemble where we know the forced signal (as the ensemble-mean trend). However, for observations or single model realizations, we must estimate the forced signal, lest it projects onto the dynamical adjustment calculations. We present several different options: a linear trend, a quadratic

trend, a high-pass filter, and a global-mean temperature trend. However, results from all options results in high RMSE and bias values. This suggests that the methods are removing part of the forced signal in addition to the internal variability, making it a poor choice for use in observations.

The lower RMSE values from the temperature MCA method also promotes the question of using only the temperature field to separate out the contribution from internal variability and that from radiative forcing. An early example in the literature is the cold-ocean-warm-land (COWL) pattern, a spatial pattern that dominates the month-to-month temperature variability in the Northern Hemisphere. Specifically, when oceans are anomalously cold, the land continents are anomalously warm. This is thought to be due to atmospheric dynamics and the contrast in thermal inertia between continents and oceans, and thus by removing this pattern we can isolate the radiative signal (Wallace et al. (1995)). However, despite explaining a substantial amount of the month-to-month variance in hemispheric and global-mean temperatures, the COWL method cannot be considered to be solely dynamically-induced since greenhouse gases warms the continents more rapidly than the oceans (Broccoli et al. (1998)).

The PCR and PLS methods rely on a seemingly arbitrary decision of the number of PCs and predictors to use, respectively. Figure 3.9 shows the RMSE values from using a different number of PCs in the PCR method. As expected, the more PCs are used, the lower the RMSE values become. The choice of the number of PCs or predictors to use is a subjective decision in every situation. If we choose too many, we will explain more of the temperature variance, but risk overfitting and capturing other noise apart from internal dynamics. If we choose too few, we risk not capturing the entirety of the atmospheric circulation effects. Here we choose to use the first 10 PCs and the first 3 predictors (although we show results for both 1 and 3 predictors) as they explain a large majority of the variance in the SLP field, with the variance explained leveling off afterwards. These choices also act as a method of comparison based on previous studies (e.g., Thompson et al. (2009); Smoliak et al. (2015)). However, there are other studies which choose differently, such as Saffioti et al. (2016), who only use the first 5 principal components to calculate dynamical adjustment.

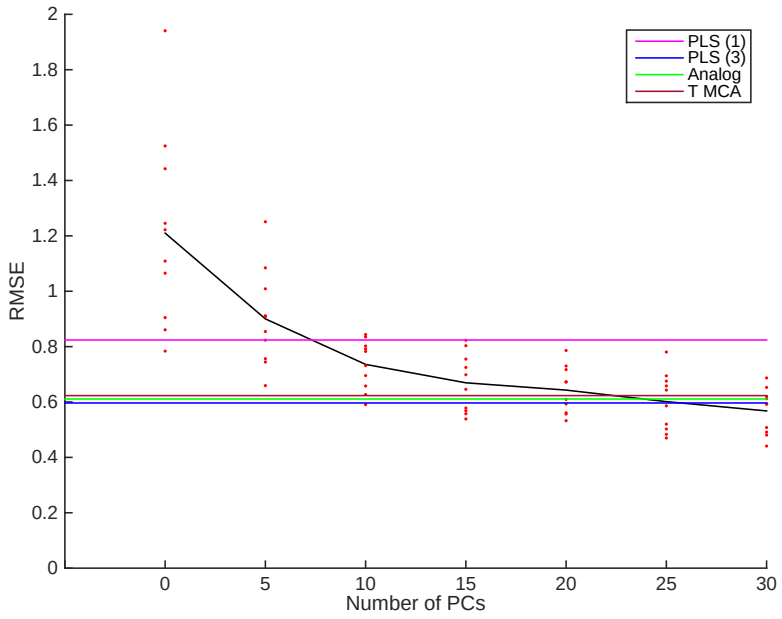

Fig. 3.9. RMSE of 50-year dynamically-adjusted temperature trends for all SLP methods in the control run. The red dots show RMSE for all 10 ensembles as a function of the number of PCs used in the PCR method, starting with zero (unadjusted) and going up to 30 PCs. The horizontal lines correspond to the mean RMSE values (left column of Table 3.1) from the other listed methods.

3.4 CONCLUSIONS

The many dynamical adjustment methods in the climate literature are motivated by the need to accurately estimate the anthropogenic signal from the noise of internal variability. However, the methods' varying complexity and different contexts naturally lead to the question "which method should I use?" The main purpose of this study is to summarize and assess the most commonly used dynamical adjustment methods and provide novel insights into the pros and cons of each method. By presenting a comprehensive analysis using both a preindustrial control run and a large ensemble of coupled model simulations, we are able to provide novel quantitative comparisons between different methods. These models allow us to examine both the uncertainty of the dynamically-adjusted temperature results and its comparisons to a true forced signal, something that would not be possible from only observational data.

Table 3.5. Advantages and drawbacks of dynamical adjustment methods.

Method	**Advantages**	**Drawbacks**
SLP-based methods	forced signal in SLP is small	not perfectly correlated with temperature advection, both responds to and forces temperature
SLP PCR, 10 PCs	simple to calculate and understand	uncertain how many PCs to use, easy to overfit
SLP MCA	simple and clear method	does not remove as much of the variance
SLP PLS	powerful method, very effective at removing more variance with few predictors	uncertain how many predictors to use, difficult to physically interpret predictors 2 and 3
SLP Constructed Analog	conceptually simple, takes advantage of control run	requires a control run, many design constraints, difficult to reproduce
T-based methods (T MCA, COWL, etc.)	better correlated with variations in the temperature field, conceptually simple	samples the forced signal in temperature, leading to inaccurate results
wind-based PLS	directly related to temperature advection, small forced signal	does not capture the full dynamical adjustment pattern

Our results show that all of the main dynamical adjustment methods show strong agreement in picking out large scale circulation-induced temperature trends. However, the attribution of smaller regional trends to internal dynamics differs amongst methods. The PLS and constructed analog methods perform best in both the preindustrial control and the large ensemble, with the lowest RMSE values to the true forced signal. However, bias analyses suggest that these methods may be bringing in the radiatively-forced response, resulting in larger biases in its dynamically-adjusted trends. As such, these two methods may not be advantageous over its counterparts in every situation.

Table 3.5 summarizes the methods compared in this paper and lists some advantages and drawbacks for each method. Smoliak et al. (2015) suggest using the first 3 predictors for the PLS method as a conservative estimate. However, we found significant biases due to the 2nd and 3rd predictors of the PLS method in the CESM-LE which worsened our estimate of the forced signal. While the use of additional predictors remove increasingly larger percentages of internal variability, they appear to also be removing portions of the externally forced signal. As such, we suggest

starting with only the first predictor of the PLS method, and only adding subsequent predictors if the pattern of variability explains a meaningful amount of variance and has a physical explanation.

4 Increased Surface Temperature Persistence in a Warming World[1]

Climate change has been and will be accompanied by widespread changes in surface temperature. It is clear these changes include global-wide increases in mean surface temperature and more provincially dependent changes in temperature variance (Hartmann et al. (2013); Kirtman et al. (2013); Collins et al. (2013)). It is less clear whether they also include changes in the day-to-day persistence of surface temperature. The current evidence is limited primarily to: 1) Simulations forced with prescribed increases in sea-surface temperatures, which have known biases in extreme events (Fischer et al. (2018)) and - by construction - do not allow changes in surface temperature persistence over maritime areas (Pfleiderer et al. (2019)), and 2) Hypothesized linkages between Arctic sea-ice declines and the midlatitude circulation (Cohen et al. (2014); Coumou et al. (2018); Francis et al. (2018)), which are not reproducible in numerous analyses and thus controversial (Barnes and Polvani (2015); Chen et al. (2016); Huguenin et al. (2020)). Here we exploit output from 1) a large-ensemble of simulations run on a coupled atmosphere/ocean general circulation model and 2) experiments run on a simplified climate model to demonstrate that climate change is associated with widespread increases in surface temperature persistence that are most pronounced over Arctic and maritime regions. It is hypothesized that the changes in surface temperature persistence are a consequence of two fundamental responses of the climate system to surface warming: 1) increases in the surface heat capacity over the Arctic and 2) reduced longwave radiative damping of surface temperature anomalies over much of the globe, especially over maritime regions. The results have important implications for the duration of extreme temperature events.

4.1 INTRODUCTION

The climate system response to increasing greenhouse gases includes changes not only in the mean temperature but also in the shape of the temperature distribution (Hartmann et al. (2013); Kirtman et al. (2013); Collins et al. (2013)). Both are essential for understanding changes in

[1]This chapter contains material that has been submitted to Nature as: Li, J. and D. W. J. Thompson: Increased surface temperature persistence in a warming world. *Nature*, submitted 10/2020.

extreme temperatures. Changes in the mean temperature alter the likelihood of extreme temperature events by shifting the probability distribution towards higher values (Hartmann et al. (2013)); changes in the temperature variance and skewness affect the incidence of extreme temperature events by altering the shape of the distribution (Lavell et al. (2012)).

The evidence for changes in the mean surface temperature is overwhelming and incontrovertible in both observations and numerical simulations of climate change (Hartmann et al. (2013); Kirtman et al. (2013); Collins et al. (2013)). The evidence for changes in the variance and skewness of the temperature distribution is less extensive. Nevertheless, it is clear that both have changed in select regions, and that these changes are physically consistent with other changes in the climate system (Schär et al. (2004b); Seneviratne et al. (2006); Donat and Alexander (2020); Fischer and Schär (2009); Fischer et al. (2012); Volodin and Yurova (2013); Lewis and Karoly (2013); McKinnon et al. (2016)). For example, Arctic amplification leads to a reduction in the meridional surface temperature gradient at high latitudes, and thus a reduction in the temperature variance due to horizontal advection (Screen (2014); Schneider et al. (2015)). Variations in precipitation and thus soil moisture lead to variations in the surface heat capacity and thus the temperature variability (Schär et al. (2004b); Seneviratne et al. (2006); Fischer and Schär (2009); Fischer et al. (2012)). Changes in extratropical dynamics lead to changes in the skewness of surface temperature (Tamarin-Brodsky et al. (2019, 2020)).

Understanding changes in the mean, variance, and skewness of surface temperature are all key for characterizing changes in the amplitude and incidence of extreme temperature events. However, they do not necessarily inform changes in the persistence of extreme events. This is important, since the persistence of an extreme temperature event plays a critical role in determining its climate impacts. Interestingly, relatively few studies have explored changes in the persistence of surface temperatures under climate change. The evidence remains limited. Pfleiderer and Coumou (2018) examine surface temperature persistence in the 1950-2014 climatology and find sporadic changes in the duration of observed extreme temperature events over the latter half of 20th century, with significant changes limited to the summer months and to land data averaged over select spatial regions. DiCecco and Gouhier (2018) infer changes in temperature persistence via changes in the

shape of temperature power spectra in climate change simulations. But their results are characterized by large geographical variations in the signs of the trends and the robustness of the changes across model simulations is not clear. Pfleiderer et al. (2019) argue that surface temperatures exhibit increased persistence in atmospheric models forced with prescribed 2K increase in sea surface temperatures, partly due to a weakening of the storm tracks. But such experiments have known biases in the attribution of extreme events (e.g., Fischer et al. (2018)), and the most robust increases are limited to select regions of the Northern Hemisphere land areas during summer. Numerous studies have argued that Arctic sea ice loss leads to systematic changes in the midlatitude circulation, including its persistence (Cohen et al. (2014); Coumou et al. (2018); Francis et al. (2018)). But the hypothesized linkages are not reproducible in many observational and modeling studies and thus controversial (Barnes and Polvani (2015); Chen et al. (2016); Huguenin et al. (2020)).

Here we exploit output from 1) a large ensemble of simulations run on a comprehensive coupled atmosphere/ocean general circulation model supplemented with 2) experiments on a simplified climate model to demonstrate that surface temperature persistence increases robustly in response to climate change across much of the globe. In contrast to previous studies, the most pronounced increases are found not over land areas, but rather over maritime areas and the Arctic. It is argued that the changes in surface temperature persistence are consistent with two fundamental responses of the climate system to global warming: 1) increases in the surface heat capacity over the Arctic as the areal coverage of sea-ice decreases and 2) reduced longwave radiative damping of surface temperature anomalies as the longwave opacity of the atmosphere increases.

4.2 METHODS

4.2.1 Calculating the autocorrelation

The autocorrelation functions are found as follows:

1) We obtain daily-mean surface temperature output from all 40 ensemble members from two periods that represent the historical and future climate: 1970-1999 for the historical period and 2070-2099 for the future period.

2) We subtract the time-varying ensemble mean temperature from all grid points in all ensemble members over both periods. Removing the ensemble mean is essential so that the surface temperature persistence in each ensemble member is not biased by low-frequency variability in the forcings applied to the model, including the increases in atmospheric carbon dioxide.

3) We estimate the persistence of surface temperature from the autocorrelation of the temperature time series (see Chapter 2). The autocorrelation is calculated as:

$$\alpha_n = \frac{\overline{T'(t)T'(t+\tau)}}{\overline{T'(t)^2}} \tag{4.1}$$

where T denotes surface temperature, primes denote departures from the ensemble mean, overbars denote the time mean, and τ denotes the lag autocorrelation in days.

4) Ensemble and zonal mean autocorrelations are calculated using the Fisher Z-transformation. That is, we transform autocorrelations to their respective z-values, average of difference the resulting z-values, and then back-transform the mean z-values to obtain the resulting mean or difference in correlations.

4.2.2 Statistical significance of autocorrelation changes

The significance of the differences in persistence is assessed in two ways:

1) Requiring at least 30 out of 40 ensemble members to agree on the sign of the change. If there is no change in temperature persistence (the null hypothesis), then we expect an equal chance of either an increase or a decrease in the temperature autocorrelation values at a single grid box in a single ensemble member between the two periods. The likelihood of at least 30 out of 40 ensemble members exhibiting the same sign of the change (either positive or negative) can be modeled using a binomial distribution with a probability of success of 0.5. That is:

$$P(x \geq 30) = \sum_{30}^{40} \binom{40}{x} (0.5)^x (0.5)^{40-x} \sim 0.1\% \tag{4.2}$$

2) Requiring that the ensemble mean lag-five autocorrelation changes exceed the 99% confidence threshold. The significance is calculated using a comparison of means test as follows:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} \tag{4.3}$$

where x_1 and x_2 are the 40 ensemble members from the 1970-1999 and 2070-2099 periods. The degrees of freedom used for the t-score is 78 (40 + 40 - 2). Note that this is a very conservative test, as this treats the autocorrelation values from each ensemble member as an individual grid point (rather than comparing two correlations using the length of the underlying data used to calculate the correlations). The results are also controlled for the false discovery rate Wilks (2016).

For a closer look at the differences between these two methods applied to SAT persistence changes, see Figure B4.

4.2.3 Spectral analyses

The spectra in Figure 4.5 are found by 1) computing the power spectrum of surface temperature at each grid box on the globe, 2) averaging the power spectra over all grid points corresponding to land areas (left), ocean areas (middle), and the Arctic (right), and 3) averaging the spatial-mean power spectra over the 40 ensemble members in the historical and future simulations. The spectra are found by applying a Hamming window and calculating the FFT of the time series. The time series are 365*30 in length.

The areas under the spectra in the top row are normalized such that $\sum_{f=0}^{F_N} S(f)\Delta f = \sigma^2$, where $S(f)$ is the power spectral density, f is the frequency, Δf is the frequency bandwidth, F_N is the Nyquist frequency, and σ^2 is the variance. The areas under the spectra in the bottom row are normalized so that $\sum_{f=0}^{F_N} S(f)\Delta f = 1$.

4.3 RESULTS

4.3.1 An Overview of temperature persistence changes in the CESM-LE

We use daily-mean surface temperature output from the CESM-LE model (Kay et al. (2015)). The 40 climate change simulations are forced with historical forcings until 2005 and the IPCC Representative Concentration Pathway 8.5 scenario forcings (RCP8.5) thereafter. They differ only in their initial conditions. As such, the results are drawn from 1) a numerical framework that allows sea-surface temperature persistence to change (i.e., in contrast to the previous +2K experiments in Pfleiderer et al. (2019)) and 2) a very large sample size that allows us to isolate the signatures of forced climate change from internal variability (Deser et al. (2012b)). The ensemble mean temperature is removed from all time series in the analyses. This removes the warming trends and thus minimizes the effects of external forcings on persistence.

Surface temperature persistence is estimated from the autocorrelation function of daily-mean temperatures; differences in persistence are estimated as the percentage change in variance explained by the autocorrelation functions. Unlike previous studies (e.g., Pfleiderer and Coumou (2018); Pfleiderer et al. (2019)) which defined persistence as as the amount of time a variable exceeds a threshold (for example, as the number of consecutive warm or cold days), the autocorrelation function takes into account the entire temperature timeseries and provides an overall representation of how temperature behaves on subseasonal timescales. The effects of climate change on temperature persistence are assessed by comparing output from two 30-year periods that represent the historical and future climate: 1970-1999 and 2070-2099, respectively. Results from the historical period in the CESM-LE are shown to be consistent with those from observations (see Figure B3).

Figure 4.1 examines the autocorrelation values and changes for a lag of 5 days in both the winter (December-January, DJF) and summer (June-August, JJA) seasons. Panels (a) and (d) show the 1970-1999 ensemble-averaged autocorrelation values at a lag of 5 days. We highlight results at a lag of five days but the patterns are largely insensitive to the specific time lag (i.e., very similar patterns are derived using time lags between 1 and 30 days). Panels (b) and (e) show the differences

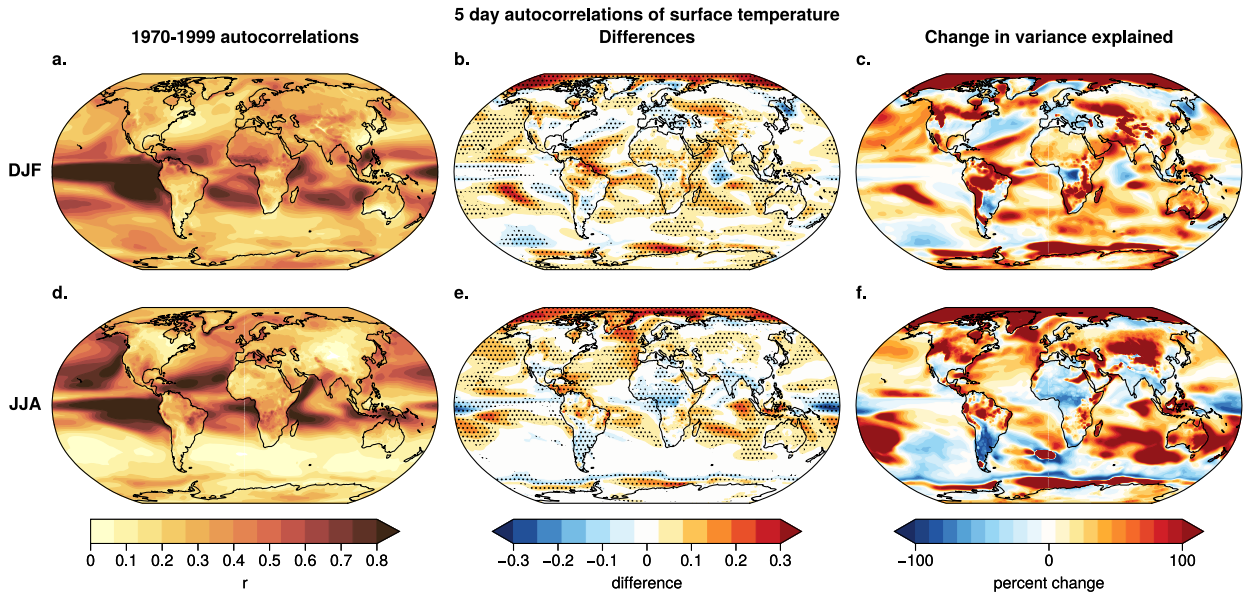

Fig. 4.1. (a,d) Surface temperature autocorrelation at a lag of 5 days in the CESM ensemble mean for DJF and JJA months. (b,e) Difference between 5 day autocorrelations in the historical (1970-1999) and future (2070-2099) periods. (c,f) Percent change in the lag-5 day autocorrelation squared (the variance explained) between the two periods. The percentage change in the variances explained is found as $\frac{r(\tau)_F^2 - r(\tau)_H^2}{r(\tau)_H^2}$, where F and H denote the future and historical periods, respectively. Warm (cool) colors represent an increase (decrease) in the five day autocorrelation in 2070-2099. Stippling indicates grid points that are significant under both tests (see Methods).

in autocorrelations between the historical and future climate change periods (2070-2099 period minus 1970-1999 period). Stippling represents grid points that satisfy both conditions of statistical significance (see 4.2). Overall, most of the globe show an increase in temperature persistence over both seasons, with large regional increases over the Arctic and select ocean areas. However, raw differences do not account for the relative change to the historical autocorrelations. As such, we also assess the percent change in variance explained by the 5 day autocorrelations in panels (c) and (f).

While there are some seasonal differences in both the autocorrelations and their changes, the winter and summer results are not significantly different from one another. To assess results for the entire temperature timeseries, Figure 4.2 shows autocorrelations based on output for all calendar days. When compared to Figure 4.1, we retain the major spatial patterns in both the autocorrelations and their percent changes in both seasons. Surface temperature persistence in the historical

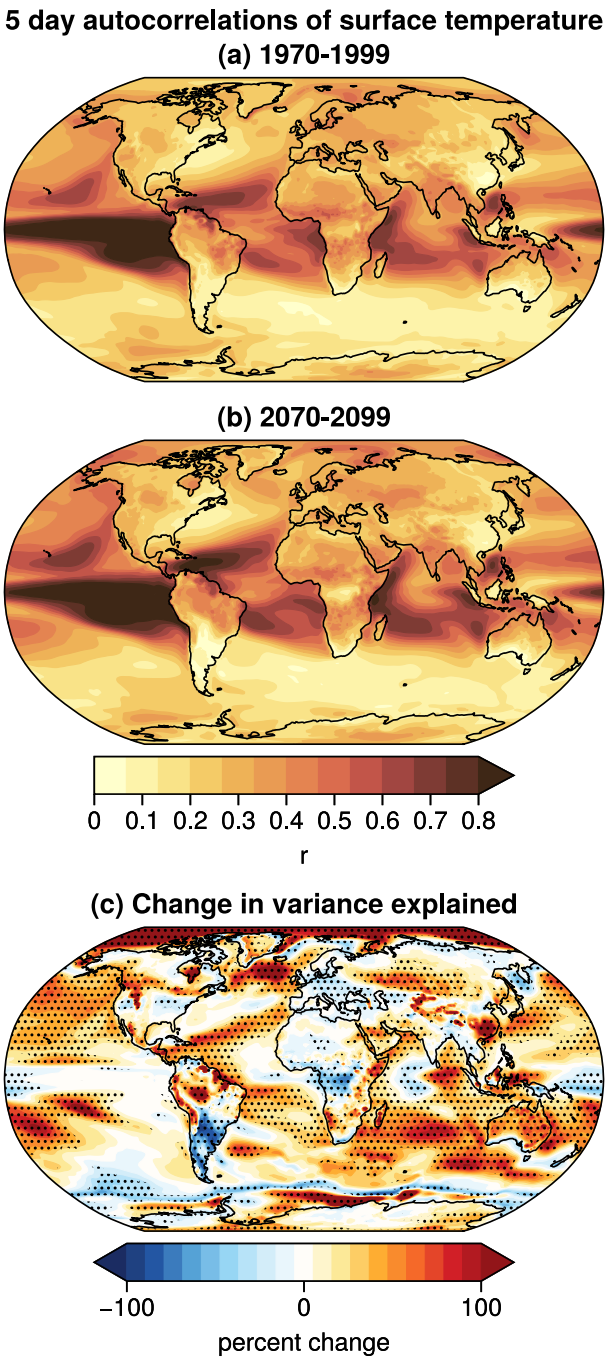

Fig. 4.2. Surface temperature autocorrelation at a lag of 5 days in the CESM ensemble mean for (a) the 1970-1999 historical period; (b) the 2070-2099 future period; and (c) the percent change in the lag-5 day autocorrelation squared (the variance explained) between the two periods, as in Figure 4.1c,f. Warm (cool) colors represent an increase (decrease) in the five day autocorrelation in 2070-2099. Stippling indicates grid points that are significant under both tests (see Methods).

period is characterized by broad maxima extending across the tropics (Figure 4.2a). The region of enhanced persistence in the eastern tropical Pacific is consistent with the model's El-Niño/Southern Oscillation (ENSO; Bulgin et al. (2020)). The maxima in other regions of the tropics are consistent with both the tropics-wide response to ENSO (Yulaeva and Wallace (1994)) and the fact that tropical surface temperatures do not vary much from one day to the next due to the small horizontal gradients in temperature there (Charney (1963)).

At first glance, there appears to be little difference between surface temperature persistence in the historical (Figure 4.2a) and future (Figure 4.2b) periods. But the percent changes in variance explained tell a different story (Figure 4.2c). The future scenario is marked by widespread increases in persistence that exceed 50% over much of the globe (Figures 4.1c,f, 4.2c). By far the most pronounced and robust increases in temperature persistence are found not over land areas - as suggested in previous studies (e.g., Pfleiderer et al. (2019)) - but over the Arctic and maritime regions. In fact, 61% of all grid points poleward of 65N and 45% of all ocean grid points exhibit significant increases in persistence between the historical and future scenarios. In contrast, only 35% of all land grid points exhibit significant increases, with the most robust results found over Australia. The only notable exceptions are found in the equatorial Pacific and high latitude South Pacific Ocean (Figure 4.2c). The increases in persistence over the Arctic are consistent with observed increases in the duration of Arctic warm events (Graham et al. (2013)). To our knowledge, the widespread increases in persistence over much of the global oceans has not been appreciated in previous work.

4.3.2 Investigating regional differences in temperature persistence changes

Figure 4.3 shows the percent changes in the autocorrelation functions averaged over all longitudes as a function of lag and stratified by land and ocean areas. Results are shown for land and ocean grid points, land grid points only, and ocean grid points only. Broadly speaking, the increases in temperature persistence are roughly uniform with lag, which supports our use of the autocorrelation as a simple measure of the changes in surface temperature memory. As also evidenced in Fig. 4.2c, the largest increases in persistence are found in the Arctic region and over

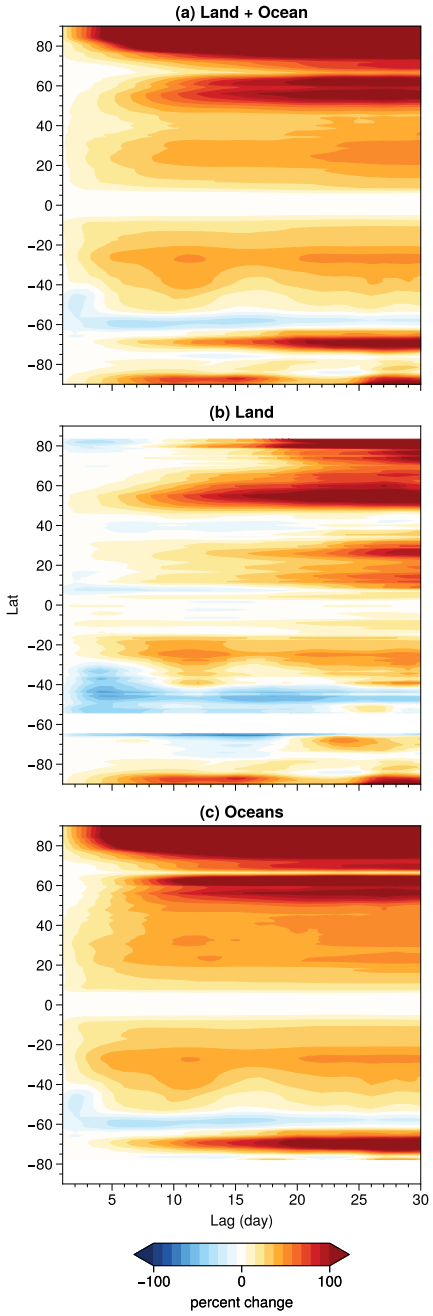

Fig. 4.3. Zonal-mean percent changes in variance explained by the lag autocorrelation between the periods 1970-1999 and 2070-2099. Warm (cool) colors represent an increase (decrease) in temperature persistence under climate change. Results are averaged over grid points corresponding to (a) land and ocean regions, (b) land regions only, and (c) ocean regions only. The results were derived by 1) computing the autocorrelation functions at all grid points in all ensemble members; 2) averaging the autocorrelation functions over all ensemble members and all longitudes for the areas indicated using the Fisher z-transformation; and 3) calculating the percent changes in variances explained by the resulting mean autocorrelation functions between the historical and future simulations.

most of the global ocean, with the exceptions of the deep tropics and high latitude Southern Ocean. The land regions exhibit relatively small changes in persistence, with some latitudes showing increased memory, but others showing little to no increases. Our results are broadly consistent with previous studies that indicate increases in surface temperature persistence over Northern Hemisphere midlatitude land regions, particularly during the summer months (Pfleiderer et al. (2019); also see Figure 4.1). But our results also highlight the relatively subtle nature of the increases in temperature persistence over land relative to those over the Arctic and maritime regions. They thus highlight the importance of assessing changes in surface temperature persistence in a coupled atmosphere/ocean climate model rather than a fixed-sea surface temperature simulation (e.g., as done in Pfleiderer et al. (2019)).

The characteristics and robustness of the changes in persistence are further explored in Figures 4.4 and 4.5. Figure 4.4 shows the autocorrelation functions for spatially-averaged temperatures (top) and the spatially-averaged autocorrelation functions for grid box temperatures (bottom) for a lag of 1 to 30 days. Results are computed for temperatures averaged over all land grid-boxes (left), all ocean grid-boxes (middle), and all Arctic grid-boxes (right). Light red lines indicate the lag-autocorrelations for all 40 ensemble members during the 2070-2099 period; light blue lines the lag-autocorrelations for all ensemble members during the 1970-1999 period; heavy lines indicate the corresponding ensemble mean values. The results in the top row provide a sense of the changes in persistence on large spatial scales, in which much of the noise due to stochastic fluctuations in regional weather is averaged out of the analysis. The bottom row gives a sense of the changes in persistence on small spatial scales, in which the day-to-day persistence is strongly influenced by weather noise. As expected, the increases in persistence are much more pronounced on large-spatial scales. The increases in persistence are most pronounced in the Arctic and global-mean ocean time series, but are interestingly also robust in the global-mean land time series. The remarkable robustness of the results is evidenced by the fact that the differences are apparent in virtually every ensemble member. The increases in persistence are less pronounced on spatial scales of 1x1 degree, but are nevertheless readily apparent over the Arctic and ocean areas, with very little overlap between the two periods' ensemble distributions, and distinct differences in the

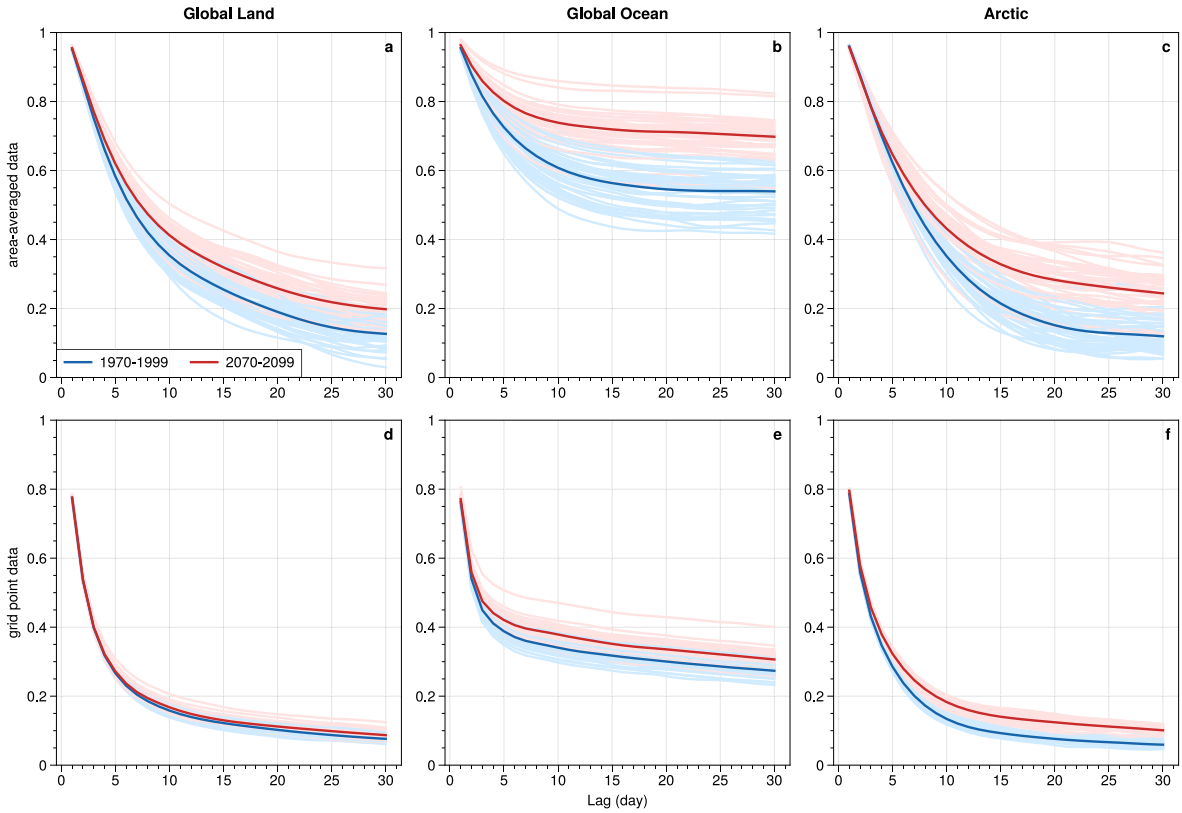

Fig. 4.4. Autocorrelation functions for spatially-averaged and grid-point temperatures. (top) The autocorrelation functions for temperatures averaged over the indicated regions. (bottom) As in the top panel, but for the autocorrelation functions computed first at each 1 deg. x 1 deg. grid point (roughly 110km by 80km at 45 degrees latitude) and then averaged via the Fisher z-transformation over the same spatial regions used to construct the top row. Light red lines indicate the lag-autocorrelations for all 40 ensemble members during the 2070-2099 period; light blue lines the lag-autocorrelations for all ensemble members during the 1970-1999 period; heavy lines indicate the corresponding ensemble mean values.

ensemble-mean, globally-averaged results. As noted earlier, the differences over the land areas are more modest, but even in the grid point power spectrum the ensemble-averaged autocorrelation curve from the future scenario is greater than that from the historical simulation at all lags.

It is important to emphasize that the ensemble mean temperature has been removed from all time series. Thus the changes in persistence highlighted in Figures 4.1 - 4.4 do not reflect changes in the timescales of the forcings between the historical and future simulations. That is, they do not reflect differences in the rates of warming due to differences in the rates of carbon dioxide increases in the two simulations. Rather, they reflect the influence of the forcings on the time-varying behavior of surface temperature about the ensemble-mean.

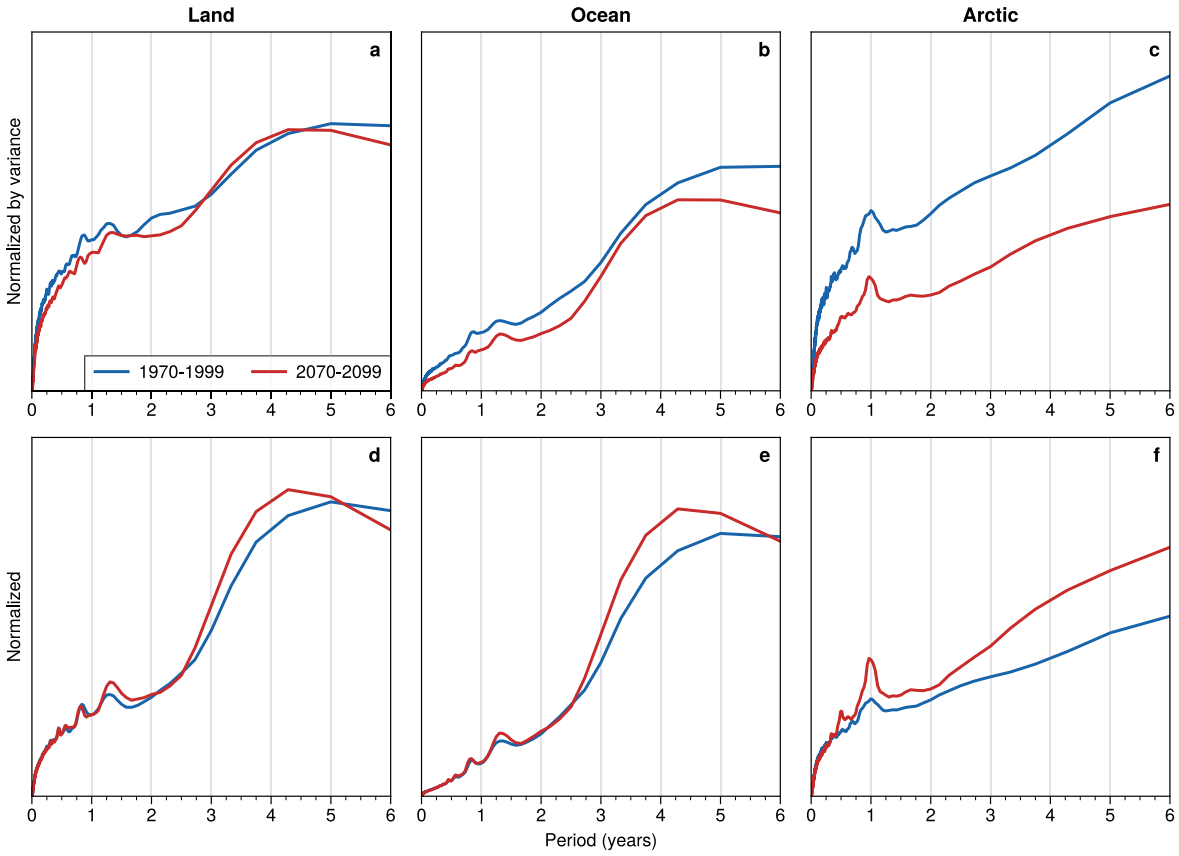

Fig. 4.5. Grid point power spectra averaged over indicated regions. Surface temperature power spectra are calculated individually at all grid points and then averaged over regions corresponding to land areas (left), ocean areas (middle), and the Arctic (right). Blue and red indicate the ensemble means of the power spectra averaged over the historical and future simulations, respectively. (top) the area integrated under the curves is proportional to the total variance of the temperature time series; (bottom) the area integrated under the curves is normalized to unit variance. The power spectra are smoothed using a 5 point running mean.

Figure 4.5 shows the power spectra of surface temperature averaged over all land, ocean, and Arctic grid points. The spectra in the top row are normalized so that the areas under the curves are proportional to the total variances of the temperature time series; the spectra in the bottom row are normalized to unit variance. As such, the areas under the red and blue curves are 1) proportional to the variance of the historical and future time series in the top row, but 2) identical in the bottom row. Note that the spectra are derived from an enormous number of independent samples, since spectra are formed at all grid points and ensemble members before being averaged.

Consider results for the Arctic. Since the areas under the curves in the top panel are proportional to the variance of the time series, Fig. 4.5c indicates that surface temperature variance

on the scale of individual grid points has on average decreased between the historical and future simulations. This is also confirmed when examining the variance of temperature anomalies in the CESM-LE (Figure 4.6). But since the areas under the curves in the bottom row are identical, Fig. 4.5f makes clear that the shape of the power spectrum has changed, with an increasingly large fraction of the variance in Arctic surface temperatures found at low frequencies. Note that the fractional increases in power at low frequencies in Fig. 4.5f are opposed by fractional decreases in power at high frequencies, but that the decreases at high frequencies are obscured by the scaling used in the figure (i.e., the area is found by integrating over frequency, but the results are shown as a function of period; see Methods). The critical point is that whereas the variance of Arctic temperatures decreases in the future simulation, the persistence of Arctic temperatures increases, as evidenced by the reddening of grid point power spectra averaged over the Arctic region.

Results for power spectra averaged over the ocean and land areas tell similar stories. In general, the variance of ocean temperatures and land temperatures decreases when averaged over all grid points (the changes in land temperature are largest for periods less than a year). However, an increasing fraction of the variance of ocean and land temperatures is found at low frequencies. Changes in land surface temperature variance (Figure 4.6) are discussed in previous studies (Schär et al. (2004b); Seneviratne et al. (2006); Donat and Alexander (2020); Fischer and Schär (2009); Fischer et al. (2012); Volodin and Yurova (2013); Lewis and Karoly (2013); Screen (2014); Schneider et al. (2015); McKinnon et al. (2016)) and are not the focus of this study. But the robust reddening of the power spectra - especially over maritime and Arctic regions - reveals a seemingly fundamental aspect of global warming: widespread increases in surface temperature persistence, particularly over maritime areas.

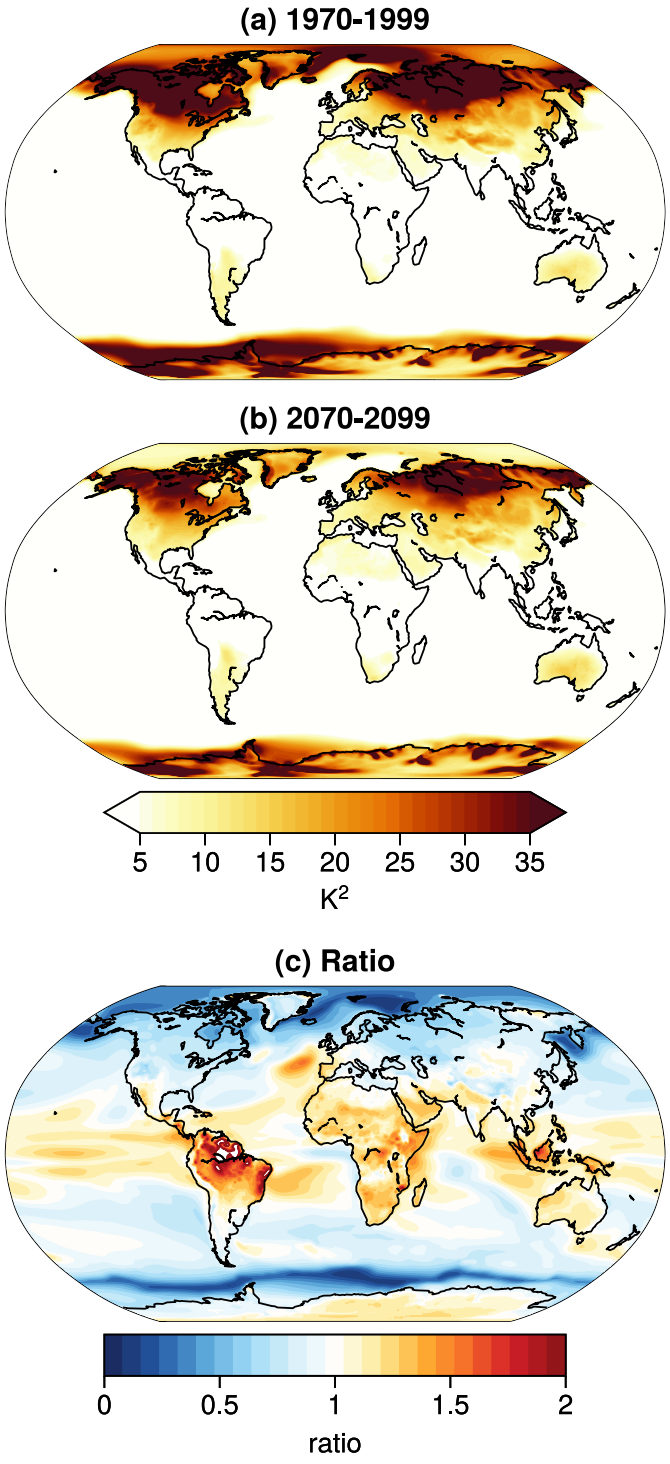

Fig. 4.6. Variance on daily surface temperature in the CESM ensemble mean for (a) the 1970-1999 historical period; (b) the 2070-2099 future period; (c) the ratio of the variances between the two periods. Warm (cool) colors in panel (c) indicate an increase (decrease) in the variance in the 2070-2099 period. Variance is calculated on temperature anomalies (the ensemble mean temperature is removed in all members) over all calendar days.

4.4 DISCUSSION

Why does surface temperature persistence increase between the historical and future simulations? Insight into the processes that drive changes in persistence can be gleaned through a simple stochastic model of the surface energy budget:

$$C_{eff}\frac{dT_s'}{dt} = F' - \lambda T_s' \tag{4.4}$$

where T_s' denotes the surface temperature anomaly, C_{eff} the effective heat capacity of the ocean-mixed layer and/or land surface, F' the stochastic forcing of surface temperature by internal variability, and a linear damping coefficient (Frankignoul and Hasselman (1977)). The damping coefficient may be viewed as reflecting the damping of surface temperature anomalies by the surface fluxes of latent, sensible and radiant heat.

There are three ways to change the persistence of T_s'. One is to change the character of the forcing of surface temperature by internal climate dynamics, F'. This line of reasoning has been used to argue that increased surface temperature persistence under climate change arises from the increased persistence of the summertime storm track (Pfleiderer et al. (2019)). However, changes in the midlatitude circulation varies from one climate change simulation to the next (Barnes and Polvani (2013)), and they cannot account for changes in persistence beyond the storm track regions. We argue that a simpler and more robust way to increase the persistence of surface temperature anomalies under climate change is via the surface heat capacity and - in particular - the effectiveness of damping by radiative processes.

If we assume that the stochastic forcing of surface temperature by internal climate variability can be characterized by white noise, then the autocorrelation of T_s' in Equation 4.4 is an exponential function with an e-folding timescale of C_{eff}/λ. Hence the ratios of the e-folding timescales between future and historical climate states are given by:

$$\frac{T_e^F}{T_e^H} = \frac{C_{eff}^F}{C_{eff}^H}\frac{\lambda^H}{\lambda^F} \tag{4.5}$$

where T_e denotes the e-folding timescale of T_s', and superscripts F and H denote the historical and future states, respectively. From Equation 4.5 it follows that increases in the surface heat capacity and/or decreases in the linear damping coefficient will lead to increases in the persistence of surface temperature.

Changes in the heat capacity are likely critical for changes in surface temperature persistence over the Arctic. The effective heat capacity of the ocean mixed-layer is at least an order of magnitude larger than that of ice. As large regions of the Arctic are transformed from ice to open water under climate change, the heat capacity of the Arctic surface must increase dramatically. The effective heat capacity of the land surface is also predicted to change in response to climate change, and soil moisture-temperature feedbacks are predicted to lead to differences in surface temperature variance over certain regions (Seneviratne et al. (2006)). But the influence of land surface changes on the time-varying characteristics of surface temperature variability is likely to vary from one region to the next and depend on model representations of land-surface processes.

A potentially key - but seemingly unappreciated - way to change surface temperature persistence is via the longwave damping of surface temperature anomalies. Climate change is marked by increases in the optical thickness of the atmosphere due to increases in both water vapor and anthropogenic emissions of carbon dioxide and other well-mixed greenhouse gases. As the optical depth of the atmosphere increases, the damping of surface temperature anomalies by longwave radiative processes should decrease, and - according to Equation 4.4 - the persistence of surface temperature should increase.

Consider the case of a warm surface temperature anomaly applied in the historical climate. The damping of the anomaly by the anomalous upward flux of longwave radiation will be attenuated by the anomalous downward flux of longwave radiation that arises as the lower atmosphere warms (Zeppetello et al. (2019)). The coupling between the upwelling and downwelling fluxes of longwave radiation at the surface increases the persistence of surface temperature anomalies for the same reasons that thermal coupling increases a) the persistence of midlatitude sea-surface temperature anomalies (Barsugli and Battisti (1998)) and b) the tropospheric relaxation timescale to radiative-convective equilibrium (Cronin and Emanuel (2013)).

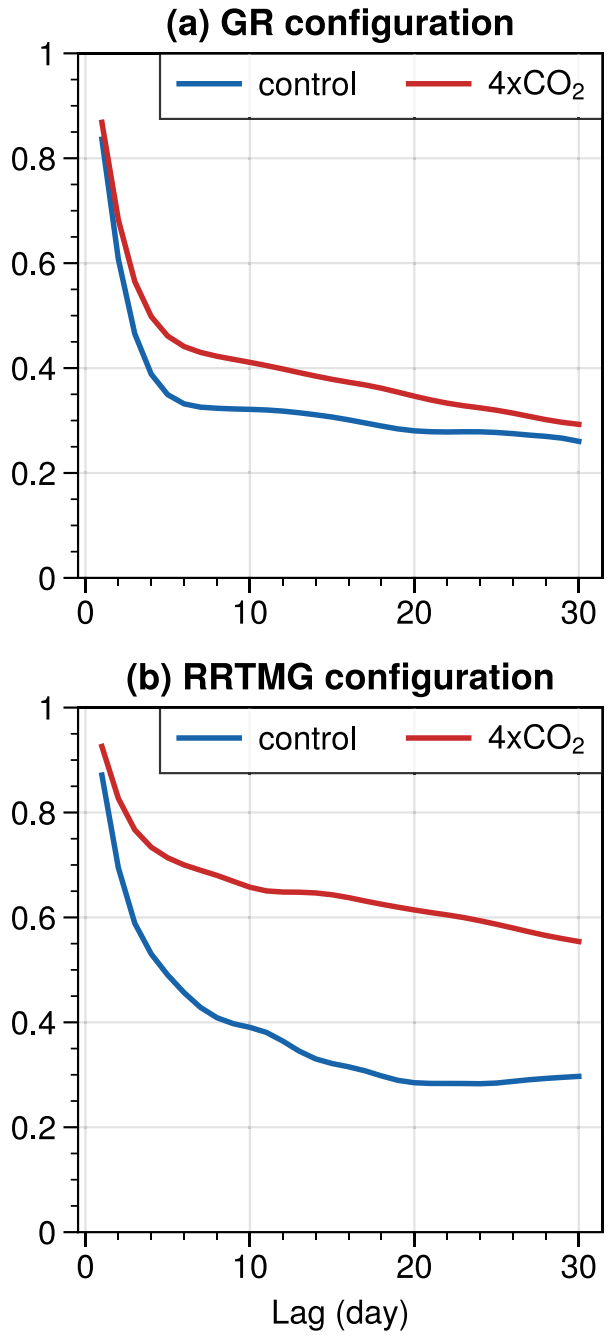

Fig. 4.7. Autocorrelation functions averaged over all grid points from two idealized numerical simulations. (top) Results from the gray-radiation atmospheric model with prescribed long wave optical depth (GR). (bottom) Results from the atmospheric model with comprehensive radiation (RRTMG). Blue curves show results for a control simulation; red curves show results from a climate change simulation in which optical depth (top) and greenhouse gas concentrations (bottom) are changed in a manner consistent with a four-fold increase in carbon dioxide. Results are from 2 different aquaplanet simulations Tan et al. (2019) and provided courtesy of Z. Tan. See text and Methods for details.

Now consider the case where the same surface temperature anomaly is applied in a future, warmer climate. In this case, the longwave opacity of the atmosphere will be higher due to increased concentrations of both well-mixed greenhouse gases and water vapor. The surface will thus be radiatively coupled to a lower and warmer level of the atmosphere, and the radiative damping of the anomaly will be less efficient than it was in the earlier, cooler climate state. That is: the radiative damping of surface temperature anomalies will be less efficient as the climate warms and the optical thickness of the atmosphere increases. In the context of the linear model (Equation 4.4), this is equivalent to a decrease in the amplitude of the linear damping of surface temperature anomalies (λ) under climate change. The same physics have been invoked to explain the differences in low-frequency surface temperature variability in numerical simulations run with and without the radiative effects of water vapor (Hall and Manabe (1999)).

Support for our hypothesis is provided by two sets of climate change experiments run on an aquaplanet general circulation model. The experiments are from the “longwave hierarchy" of numerical model configurations (Tan et al. (2019); output provided courtesy of Z. Tan). Both sets of experiments include a “control" and a “climate change" simulation run on an atmospheric general circulation model coupled to a slab-ocean. The differences between the sets lies in their treatment of atmospheric radiation: the first set is run on a “gray radiation (GR)" configuration of the model, in which the radiative effects of atmospheric composition are parameterized using a gray radiation scheme with prescribed optical depth. The second set is run on an atmospheric model with a comprehensive radiation scheme (RRTMG), in which the radiative effects of atmospheric composition are explicitly calculated by the radiation scheme. Additional details of the numerical models and simulations shown here are provided in Methods.

The control and climate change experiments are forced as follows. In the case of the RRTMG configuration: the control simulation is forced with CO_2 concentrations of 355 ppmv, and the climate-change simulation is forced with a four-fold increase in CO_2 concentrations relative to the control. In the case of the GR configuration: the simulations are forced with prescribed optical

depths that are configured so that the long-term mean sea-surface temperature fields in the GR control and climate change simulations closely match those in the corresponding RRTMG simulations. In all simulations, the ocean mixed-layer depth is fixed at 30m.

Figure 4.7 shows the autocorrelation functions for near-surface temperatures averaged over all grid points in both simulations. The globally averaged autocorrelation functions are found in the same manner as the lower panels of Figure 4.4. The figure highlights two key results:

1) *Increasing atmospheric optical depth (either explicitly as in the GR simulations or by forcing the model with increases in atmospheric greenhouse gases as in the RRTMG simulations) leads to pronounced increases in the persistence of surface temperatures.* The differences cannot be attributed to changes in the surface heat capacity, since there is no land surface or sea-ice, and the ocean mixed-layer depth is fixed.

2) *The increases in persistence are more pronounced in the numerical configuration that includes an interactive water vapor feedback (the RRTMG simulation).* The radiative effects of water vapor are included in the GR control and climate simulations only through the prescribed optical depths. Since the optical depths are fixed in time, model-derived variations in water vapor do not contribute to variations in the radiative fluxes, and thus the water vapor feedback does not contribute to the model's internal climate variability. In contrast, model-derived variations in water vapor are explicitly accounted for in the RRTMG radiation scheme. The larger persistences in the RRTMG configurations, and the larger differences between the RRTMG control and climate change simulations, are both consistent with the role of the water vapor feedback in low-frequency climate variability (Hall and Manabe (1999)).

4.5 CONCLUSIONS

Climate change simulations run on a coupled atmosphere/ocean general circulation large ensemble model reveal widespread increases in surface temperature persistence under climate change. The results shown here provide the most robust evidence to-date of widespread increases in surface temperature persistence under climate change, with by far the most pronounced increases arising over the Arctic and maritime regions.

We have argued on the basis of theory and output from climate change experiments run on a simpified general circulation model that the increases in persistence reflect a robust thermodynamic and radiative constraint on surface temperature variability. As temperatures increase, so does the longwave opacity of the atmosphere due to increases in both carbon dioxide and water vapor concentrations. The increased optical thickness of the atmosphere reduces the thermal damping of surface temperature anomalies by longwave radiative processes for the same reasons that atmosphere/ocean thermal coupling increases the timescales of sea-surface temperature anomalies (Barsugli and Battisti (1998)). The large increases in Arctic surface temperature persistence likely also reflect the increases in surface heat capacity that must occur as vast regions of the Arctic are converted from sea-ice to open water.

5 Conclusions

The work detailed in this dissertation examines surface temperature variability and its response to external radiative warming. By utilizing a large ensemble model (the NCAR CESM-LE), we are able to provide new insights into the role of internal variability and the forced response in the surface temperature field. Our research also explores changes in temperature variability itself as a response to increased radiative warming. In this concluding chapter, we summarize the main results of this dissertation and discuss possible avenues of future research.

5.1 DYNAMICAL CONTRIBUTIONS TO SURFACE TEMPERATURE TRENDS

The first part of this dissertation examined the role of internal dynamics on surface temperature trends. The atmospheric circulation plays a large role in determining surface temperature on daily to annual timescales. Many studies have examined this relationship using a variety of methods to better understand the relative roles of internal variability and external forcing in regional climate trends, or to better estimate and detect the forced signal itself. Here we explore how best to remove the internal dynamical contribution of surface temperature to better pinpoint and understand the radiatively forced signal. Several dynamical adjustment methods are applied to both a pre-industrial control run (where there are no external forcings) and a large ensemble simulation with historical forcings (the 1955-2004 period). The use of a control run tests the accuracy of the forced signal after dynamical adjustment, while the large ensemble tests the reliability of the methods under external forcing. Dynamical adjustment narrows the spread in the large ensemble and on average brings temperature trends closer to the radiatively forced signal of the ensemble average.

This study builds directly off of previous work on applications of dynamical adjustment to temperature variability and trends in both observations and modeling studies (e.g., Thompson et al. (2009); Smoliak et al. (2015); Deser et al. (2016)). We compare and assess these dynamical adjustment methods on the basis of 2 attributes: the method should remove a maximum amount

of internal variability while preserving the true forced signal. To provide a quantitative analysis between the different methods, we calculate the variance explained in the temperature trends from each method, as well as the RMSE values of the dynamically-adjusted trends from the true forced signal. Previous methods were almost exclusively based on the SLP field (Merrifield et al. (2017) based their analyses on Z500), and here we also compare regression methods using the 950hPa wind fields. While the control run is excellent for assessing the methods in an “ideal" environment, results from the CESM-LE show biases in the dynamically-adjusted trends due to a forced response in the circulation fields themselves. This is consistent with Deser et al. (2016), who partition both the forced and internal response into respective dynamic and thermodynamic components.

Our work provides a template from which to assess the various dynamical adjustment methods available to the community. Table 3.5 outlines the major advantages and disadvantages for each type of method based on our analyses. We do not find there to be an unequivocally “best" method; specific methods are more or less appropriate depending on the individual situation, research question(s), and datasets available. Moreover, recent research has introduced additional methods and tools for dynamical adjustment (e.g., Sippel et al. (2019); Wills et al. (2020)), and it is likely that new methods will continue to be added in the future. We hope to clarify several advantages and issues to help readers gain a clearer understanding to determine which method(s) is best applied to their own situation.

5.2 EXAMINING THE EFFECTS OF CLIMATE CHANGE ON TEMPERATURE PERSISTENCE

The temperature persistence response to climate change is investigated in the CESM-LE under a strong external warming scenario. We examine daily temperatures on subseasonal timescales for both seasonal and all calendar days. Comparing the 1970-1999 historical period with projections for the 2070-2099 future period reveals a broad increase in surface temperature persistence globally, with the largest changes found over the Arctic and ocean regions. While there are more

specific regional changes, we focus on the large scale changes, specifically changes over land compared to ocean grid points.

We examine temperature persistence changes both on large spatial scales (where regional weather noise is averaged out) and small grid point scales (where the persistence is averaged over the entire region; Figure 4.4). We find increases in both spatial scales for land regions, ocean regions, and the Arctic. Temperature autocorrelations over the larger spatial scales exhibit much larger variability amongst ensemble members, but also exhibit more robust increases in persistence, with very little overlap between the ensemble member spread. Increases in temperature persistence is greatest over the oceans and the Arctic, and is corroborated using spectral analysis. The increase in persistence is found alongside a decrease in temperature variance over the spatially-averaged regions.

We suggest that the almost ubiquitous increase in temperature persistence is due to changes in the surface heat capacity (especially over the Arctic) and an increase in radiative damping. This hypothesis is based off a simple model of the surface energy budget found in Frankignoul and Hasselman (1977). From this model, the e-folding timescale of temperature autocorrelation scales with the ratio between the heat capacity and the damping coefficient. In a warmer climate, the atmosphere will have higher concentrations of both greenhouse gases and water vapor, increasing the longwave opacity. Results from two aquaplanet runs with different radiation schemes support this hypothesis. Our results have important implications for extreme weather events in future climate.

5.3 THE UTILITY OF LARGE ENSEMBLES IN UNDERSTANDING CLIMATE VARIABILITY

While we used several different models and tools throughout this dissertation, the results from Chapters 3 and 4 are in large part due to the unique modeling framework provided by initial-condition large ensemble models. The internal variability of the Earth's climate system is important not only for its effects on surface weather, but also provides an inherent limit to climate predictability. Large ensemble models create many different realizations within the same climate

model under the same external forcings. This allows us to both quantify the internal variability and extract the forced response from the ensemble members.

Previous studies of both dynamical adjustment (e.g., Thompson et al. (2009); Smoliak et al. (2015)) and temperature persistence (Pfleiderer and Coumou (2018)) used only the single realization of observational data. While the results are undoubtedly useful, they are also severely limited. In the large ensemble, the external forced signal is known (provided there are enough ensemble members). This allows us to compare the dynamically-adjusted temperature trends to the true forced signal in Chapter 3. In a single realization, it is not known if the residual temperature trends are at all similar to the true forcings. In addition, anomalies are usually calculated from detrending the original time series; however, this also removes an unknown portion of the internal variability within the time series, thus also adding error to the dynamical adjustment calculations. The CESM-LE simulations allowed us to directly assess the methods within an external forcing framework, thus providing a direct comparison to analyses from the control run.

In Chapter 4, we calculate autocorrelations on temperature anomalies with the ensemble mean time series removed. Thus, the autocorrelation results are not directly due to the external forcing nor variations such as the seasonal cycle. This allows for an “apples to apples" comparison between the two time periods. A large ensemble also allowed us to evaluate the robustness of the change in temperature persistence in the context of their internal variability (e.g., Figure 4.4). We are able to show that statistical significance can be calculated solely from the ensemble members, without the need for a traditional statistics test.

We note that the results presented here are specific only to the NCAR CESM-LE model. As such, it is subject to the CESM’s model configurations, biases, interpretation of internal variability, and the external forcings applied.

5.4 FUTURE WORK

Building off the research outlined in this dissertation, we suggest several avenues for future work that would expand and deepen our understanding of the mechanisms and effects of surface temperature variability changes in future climate.

5.4.1 Forced dynamical responses to climate change

Numerous studies have examined the forced response of the circulation to external warming (e.g., Barnes and Polvani (2013); Peings et al. (2017); Coumou et al. (2018); Huguenin et al. (2020)). However, the results are far from conclusive. Results vary regionally and widely from model to model. Our bias results in Chapter 3 suggest that the SLP and 950hPa wind fields in the CESM-LE exhibit a noticeable forced response even under historical forcings, and presumably a greater response under stronger warming scenarios. Such results suggest that the uncertainty due to internal variability itself is also a response of climate change. Thus, future research distinguishing dynamical responses to external forcing (e.g., a change in the strength and location of the eddy driven jet) and their respective effects (e.g., a change in surface temperature variability due to the dynamical response) from internal variability will be invaluable in climate predictability.

One such way could be via contrastive principal component analysis (cPCA), a tool for unsupervised learning that has been used in several other fields (Abid et al. (2018)). While principal component analysis (PCA) aims to identify dominant trends in one dataset, cPCA identifies low-dimensional structures that are distinct relative to comparison data. In many cases, applying regular PCA to two datasets results in dominant variability being within the background noise; cPCA instead identifies the projections that exhibit the most interesting differences across datasets, allowing us to visualize dataset-specific patterns missed by PCA and other standard methods. For instance, the North Atlantic Oscillation (NAO) is the primary mode of variability in the North Atlantic Ocean. PCA applied to SLP in the North Atlantic for two different time periods would result in two NAO patterns that look very similar. Applying cPCA, on the other hand, will result in the differences between the two NAO patterns. Moreover, applying cPCA on the same fields in multiple models allows us to examine differences in changes between models. Preliminary research suggests that this could be a useful tool for assessing the forced dynamical response and its impacts on surface temperature variability.

5.4.2 Investigating physical mechanisms in temperature persistence changes

In Chapter 4 we demonstrate that surface temperature persistence robustly increases in the CESM-LE under radiative warming. We hypothesize that these changes are due to increased longwave radiative damping and surface heat capacity, especially over the Arctic. While results from a set of idealized models with different radiation schemes support this hypothesis, further research is needed to evaluate the relative contribution to the temperature persistence response. Both our work in Chapter 4 and previous studies have only investigated changes to persistence at the surface; examining temperature persistence changes in higher levels of the atmosphere could also provide new insights into the forced response. Previous studies have suggested that changes in the storm tracks and soil moisture contribute to increases in surface temperature persistence over regional land areas (Pfleiderer and Coumou (2018); Pfleiderer et al. (2019)). However, correlation does not equal causation. Most likely, there are several different factors that all contribute to the changes in temperature persistence, and these factors differ in strength with regionality. A modeling hierarchy can be helpful in evaluating the individual contributions to the forced response in temperature persistence; however, it must be noted that models differ in their projections of the atmospheric circulation, soil moisture and precipitation, sea ice, aerosols, and other variables. Our results in Chapter 4 only provide the initial foundation and motivation for many future studies.

5.4.3 Evaluating Persistence Changes in a Multi-Model Large Ensemble Archive

Another major source of uncertainty in climate projections is due to model error. Our results in Chapter 4 focused on examining temperature persistence changes within a single large ensemble model. While we were able to account for the uncertainty due to internal variability, we could not account for biases due to model structural differences. A way to evaluate both the internal variability and the forced response is by using multiple large ensemble models. While this has generally not been done due to the expense and difficulty of running and/or gathering large amounts of data from multiple models, a recent project assembled data from 7 different CMIP5-class large ensemble models (the Multi-Model Large Ensemble Archive (MMLEA); Deser et al. (2020)). By using the MMLEA, we can assess the robustness of our temperature persistence results from

Chapter 4 by comparing them to results from 6 other large ensemble models. Extreme events are by definition rare and thus difficult to quantify, but the large data available in the MMLEA would also allow us to assess such events (e.g., heat waves) in future climate projections. These results would have large implications for risk assessments in future projections of extreme events.

REFERENCES

Abid, A., M. J. Zhang, V. K. Bagaria, and J. Zou, 2018: Exploring patterns enriched in a dataset with contrastive principal component analysis. *Nat. Commun.*, **9**, 2134, doi: 10.1038/s41467-018-04608-8.

Barnes, E. A. and L. Polvani, 2013: Response of the midlatitude jets and of their variability to increased greenhouse gases in the CMIP5 models. *J. Climate*, **26**, 7117–7135, doi: 10.1175/JCLI-D-12-00536.1.

Barnes, E. A. and L. Polvani, 2015: CMIP5 projections of Arctic amplification, of the North American/North Atlantic circulation, and of their relationship. *J. Climate*, **28**, doi: 10.1175/JCLI-D-14-00589.1.

Barsugli, J. J. and D. S. Battisti, 1998: The Basic Effects of Atmosphere–Ocean Thermal Coupling on Midlatitude Variability. *J. Atmos. Sci.*, **55**, 477–493, doi: 10.1175/1520-0469(1998)055<0477:TBEOAO>2.0.CO;2.

Boer, G., 2009: Changes in interannual variability and decadal potential predictability under global warming. *J. Climate*, **22**, 3098–3109, doi: 10.1175/2008JCLI2835.1.

Bretherton, C. S., C. Smith, and J. M. Wallace, 1992: An intercomparison of methods for finding coupled patterns in climate data. *J. Climate*, **5**, 541–560, doi: 10.1175/1520-0442(1992)005<0541:AIOMFF>2.0.CO;2.

Broccoli, A. J., N. Lau, and M. J. Nath, 1998: The Cold Ocean-Warm Land Pattern: Model Simulation and Relevance to Climate Change Detection. *J. Climate*, **11**, 2743–2763, doi: 10.1175/1520-0442(1998)011<2743:TCOWLP>2.0.CO;2.

Bulgin, C. E., C. J. Merchant, and D. Ferreira, 2020: Tendencies, variability and persistence of sea surface temperature anomalies. *Sci. Rep.*, **10**, 7986.

Charney, J. G., 1963: A note on large-scale motions in the tropics. *J. Atmos. Sci.*, **20**, 607–609.

Chen, H. W., F. Zhang, and R. B. Alley, 2016: The robustness of midlatitude weather pattern changes due to Arctic sea ice loss. *J. Climate*, **29**, 7831–7849, doi: 10.1175/JCLI-D-16-0167.1.

Cohen, J., et al., 2014: Recent Arctic amplification and extreme mid-latitude weather. *Nat. Geosci.*, **7**, 627–637, doi: 10.1038/ngeo223.

Collins, M., et al., 2013: Long-term Climate Change: Projections, Commitments and Irreversibility. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. *Cambridge University Press*.

Compo, G. and P. D. Sardeshmukh, 2007: Oceanic influences on recent continental warming. *Clim. Dyn.*, **32**, 333–342.

Coumou, D., G. D. Capua, G. Vavrus, L. Wang, and S. Wang, 2018: The influence of Arctic amplification on mid-latitude summer circulation. *Nat. Commun.*, **9**, 2959, doi: 10.1038/s41467-018-05256-8.

Coumou, D. and A. Robinson, 2013: Historic and future increase in the global land area affected by monthly heat extremes. *Environ. Res. Lett.*, **8**, 1–6, doi: 10.1088/1748-9326/8/3/034018.

Cronin, T. W. and K. A. Emanuel, 2013: The climate time scale in the approach to radiative-convective equilibrium. *JAMES*, **5**, 843–849, doi: 10.1002/jame.20049.

Dee, D. P., et al., 2011: The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.*, **137**, 553–597, doi: 10.1002/qj.828.

Delworth, T. L., A. J. Broccoli, R. J. Stouffer, V. Balaji, and J. A. Beesley, 2006: GFDL's CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *J. Climate*, **19**, 643–674, doi: 10.1175/JCLI3629.1.

Deser, C., R. Knutti, S. Solomon, and A. S. Phillips, 2012a: Communication of the role of natural variability in future North America climate. *Nat. Clim. Change*, **2**, 775–779, doi: 10.1038/nclimate1562.

Deser, C., A. S. Phillips, M. A. Alexander, and B. V. Smoliak, 2014: Projecting North American Climate over the Next 50 Years: Uncertainty due to Internal Variability. *J. Climate*, **27**, 2271–2296, doi: 10.1175/JCLI-D-13-00451.1.

Deser, C., A. S. Phillips, V. Bourdette, and H. Teng, 2012b: Uncertainty in climate change projections: the role of internal variability. *Climate Dyn.*, **38**, 527–546, doi: 10.1007/s00382-010-0977-x.

Deser, C., L. Terray, and A. S. Phillips, 2016: Forced and Internal Components of Winter Air Temperature Trends over North America during the past 50 Years: Mechanisms and Implications. *J. Climate*, **29**, 2237–2258, doi: 10.1175/JCLI-D-15-0304.1.

Deser, C., et al., 2020: Insights from Earth system model initial-condition large ensembles and future prospects. *Nature Clim. Change*, **10**, 277–286, doi: 10.1038/s41558-020-0731-2.

DiCecco, G. J. and T. C. Gouhier, 2018: Increased spatial and temporal autocorrelation of temperature under climate change. *Sci. Rep.*, **8**, 14 850, doi: 10.1038/s41598-018-33217-0.

Donat, M. G. and L. V. Alexander, 2020: The shifting probability distribution of global daytime and night-time temperatures. *Geophys. Res. Lett.*, **39**, 14.

Fischer, E. M., U. Beyerle, C. F. Schleussner, A. D. King, and R. Knutti, 2018: Biased Estimates of Changes in Climate ExtremesFrom Prescribed SST Simulations. *Geophys. Res. Lett.*, **45**, 8500–8509, doi: 10.1029/2018GL079176.

Fischer, E. M., J. Fajczak, and C. Schär, 2012: Changes in European summer temperature variability revisited. *Geophys. Res. Lett.*, **39**, L19 702, doi: 10.1029/2012GL052730.

Fischer, E. M. and C. Schär, 2009: Future changes in daily summer temperature variability: Driving processes and role for temperature extremes. *Clim. Dyn.*, **33**, 917–935, doi: 10.1007/s00382-008-0473-8.

Francis, J. A., N. Skific, and S. J. Vavrus, 2018: North American Weather Regimes Are Becoming More Persistent: Is Arctic Amplification a Factor? *Geophys. Res. Lett.*, **45**, 11 414–11 422, doi: 10.1029/2018GL080252.

Frankignoul, C. and K. Hasselman, 1977: Stochastic climate models, Part 2. Application to sea-surface temperature variability and thermocline variability. *Tellus*, **29**, 289–305.

Frierson, D. M. W., I. M. Held, and P. Zurita-Gotor, 2006: A gray-radiation aquaplanet moist GCM. Part I: Static stability and eddy scale. *J. Atmos. Sci.*, **63**, 2548–2566, doi: 10.1175/JAS3753.1.

Garfinkel, C. I. and N. Harnik, 2017: The Non-Gaussianity and Spatial Asymmetry of Temperature Extremes Relativeto the Storm Track: The Role of Horizontal Advection. *J. Climate*, **30**, 445–464, doi: 10.1175/JCLI-D-15-0806.1.

Ghil, M., 2001: Hilbert problems for the geosciences in the 21st century. *Nonlinear Processes Geophys.*, **8**, 211–222, doi: 10.5194/ npg-8-211-2001.

Graham, R. M., L. Cohen, A. A. Petty, L. N. Boisvert, A. Rinke, S. R. Hudson, M. Nicolaus, and M. A. Granskog, 2013: Increasing frequency and duration of Arctic winter warming events. *Geophys. Res. Lett.*, **44**, 6974–6983, doi: 10.1002/2017GL073395.

Hall, A. and S. Manabe, 1999: The Role of Water Vapor Feedback in Unperturbed Climate Variability and Global Warming. *J. Climate*, **12**, 2327–2346.

Hansen, J., M. Sato, and R. Ruedy, 2012: Perception of climate change. *Proc. Natl. Acad. Sci.*, **109**, E2415–E2423, doi: 10.1073/pnas.1205276109.

Hartmann, D. L., et al., 2013: Observations: Atmosphere and Surface. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. *Cambridge University Press*.

Hawkins, E. and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions. *Bull. Amer. Meteor. Soc.*, **90**, 1095–1107, doi: 10.1175/2009BAMS2607.1.

Holmes, C. R., T. Woollings, E. Hawkins, and H. D. Vries, 2016: Robust future changes in temperature variability under greenhouse gas forcing and the relationship with thermal advection. *J. Climate*, **29**, 2221–2236, doi: 10.1175/JCLI-D-14-00735.1.

Huguenin, M. F., E. M. Fischer, S. Kotlarski, S. C. Scherrer, C. Schwierz, and R. Knutti, 2020: Lack of change in the projected frequency and persistence of atmospheric circulation types over Central Europe. *Geophys. Res. Lett.*, **47**, doi: 10.1029/2019GL086132.

Huntingford, C., P. D. Jones, V. N. Livina, T. M. Lenton, and P. M. Cox, 2013: No increase in global temperature variability despite changing regional patterns. *Nature*, **500**, 327–330, doi: 10.1038/nature12310.

Hurrell, J. W., 1996: Influence of variations in extratropical wintertime teleconnections on northern hemisphere temperature. *Geophys. Res. Lett.*, **23**, 665–668, doi: 10.1029/96GL00459.

Hurrell, J. W., et al., 2013: The Community Earth System Model: A Framework for Collaborative Research. *Bull. Amer. Meteor. Soc.*, **94**, 1339–1360, doi: 10.1175/BAMS-D-12-00121.1.

Iacono, M. J., J. S. Delamere, E. J. Mlawer, M. W. Shephard, S. A. Clough, and W. D. Collins, 2008: Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. *J. Geophys. Res.*, **113**, D13 103, doi: 10.1029/2008JD009944.

Kang, S. M., I. M. Held, D. M. W. Frierson, and M. Zhao, 2008: The response of the ITCZ to extratropical thermal forcing: Idealized slab-ocean experiments with a GCM. *J. Climate*, **21**, 3521–3532, doi: 10.1175/2007JCLI2146.1.

Katz, R. W. and B. G. Brown, 1992: Extreme events in a changing climate: Variability is more important than averages. *Climatic Change*, **21**, 289–302, doi: 10.1007/BF00139728.

Kay, J. E., et al., 2015: The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability. *Bull. Amer. Meteor. Soc.*, **96**, 1333–1349, doi: 10.1175/BAMS-D-13-00255.1.

Kharin, V. V., F. W. Zwiers, X. Zhang, and G. C. Hegerl, 2007: Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. *J. Climate*, **20**, 1419–1444, doi: 10.1175/JCLI4066.1.

Kirtman, B., et al., 2013: Near-term Climate Change: Projections and Predictability. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. *Cambridge University Press*.

Kooperman, G. J., M. S. Pritchard, and R. C. J. Somerville, 2015: The response of US summer rainfall to quadrupled CO2 climate change in conventional and superparameterized versions of the NCAR community atmosphere model. *JAMES*, **6**, 859–882, doi: 10.1002/2014MS000306.

LaJoie, E. and T. DelSole, 2016: Changes in Internal Variability due to Anthropogenic Forcing: A New Field Significance Test. *J. Climate*, **29**, 5547–5560, doi: 10.1175/JCLI-D-15-0718.1.

Lavell, A., M. Oppenheimer, C. Diop, J. Hess, R. Lempert, J. Li, R. Muir-Wood, and S. Myeong, 2012: Climate change: new dimensions in disaster risk, exposure, vulnerability, and resilience. In: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)]. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC). *Cambridge University Press*.

Lehner, F., C. Deser, and L. Terray, 2017: Towards a new estimate of "time of emergence" of anthropogenic warming: insights from dynamical adjustment and a large initial-condition model ensemble. *J. Climate*, **30**, 7739–7756, doi: 10.1175/JCLI-D-16-0792.1.

Lewandowsky, S., J. S. Risbey, and N. Oreskes, 2016: The "pause" in global warming: Turning a routine fluctuation into a problem for science. *Bull. Amer. Meteor. Soc.*, **97**, 723–733, doi: 10.1175/BAMS-D-14-00106.1.

Lewis, S. C. and D. J. Karoly, 2013: Anthropogenic contributions to Australia's record summer temperatures of 2013. *Geophys. Res. Lett.*, **40**, 3705–3709.

Li, J., D. W. J. Thompson, E. A. Barnes, and S. Solomon, 2017: Quantifying the lead time required for a linear trend to emerge from natural climate variability. *J. Climate*, **30**, 10 179–10 191, doi: 10.1175/JCLI-D-16-0280.1.

Lorenz, E. N., 1969: Atmospheric Predictability as Revealed by Naturally Occurring Analogues. *J. Atmos. Sci.*, **26**, 636–646, doi: 10.1175/ 1520-0469(1969)26,636:APARBN.2.0.CO;2.

Mantua, N. J., S. R. Hare, Y. Zhang, J. M. Wallace, and R. C. Francis, 1997: A pacific interdecadal climate oscillation with impacts on salmon production. *Bull. Amer. Meteor. Soc.*, **78**, 1069–1080, doi: 10.1175/1520-0477(1997)078<1069:APICOW>2.0.CO;2.

McKinnon, K. A., A. Poppick, E. Dunn-Sigouin, and C. Deser, 2017: An "observational large ensemble" to compare observed and modeled temperature trend uncertainty due to internal variability. *J. Climate*, **30**, 7585–7598, doi: 10.1175/JCLI-D-16-0905.1.

McKinnon, K. A., A. Rhines, M. P. Tingley, and P. Huybers, 2016: The changing shape of Northern hemisphere summer temperature distributions. *J. Geophys. Res. Atm.*, **121**, 8849–8868.

Meehl, G. A., J. M. Arblaster, and C. Tebaldi, 2005: Understanding future patterns of increased precipitation intensity in climate model simulations. *Geophys. Res. Lett.*, **32**, L18 719, doi: 10.1029/2005GL023680.

Meehl, G. A., C. Tebaldi, and D. Nychka, 2004: Changes in frost days in simulations of twentyfirst century climate. *Climate Dyn.*, **23**, 495–511, doi: 10.1007/s00382-004-0442-9.

Merrifield, A., F. Lehner, S.-P. Xie, and C. Deser, 2017: Removing circulation effects to assess central u.s. land-atmosphere interactions in the cesm large ensemble. *Geophys. Res. Lett.*, **44**, 9938–9946, doi: 10.1002/2017GL074831.

Peings, Y., J. Cattiaux, S. Vavrus, and G. Magnusdottir, 2017: Late twenty-first-century changes in the midlatitude atmospheric circulation in the cesm large ensemble. *J. Climate*, **30**, 5943–5960, doi: 10.1175/JCLI-D-16-0340.1.

Pfleiderer, P. and D. Coumou, 2018: Quantification of temperature persistence over the northern hemisphere land-area. *Climate Dyn.*, **51**, 627–637, doi: 10.1007/s00382-017-3945-x.

Pfleiderer, P., C. Schleussner, K. Kornhuber, and D. Coumou, 2019: Summer weather becomes more persistent in a 2°c world. *Nat. Clim. Change*, **9**, 666–671, doi: 10.1038/s41558-019-0555-0.

Quadrelli, R. and J. M. Wallace, 2004: A Simplified Linear Framework for Interpreting Patterns of Northern Hemisphere Wintertime Climate Variability. *J. Climate*, **46A**, 3728–3744, doi: 10.1175/1520-0442(2004)017<3728:ASLFFI>2.0.CO;2.

Rahmstorf, S. and D. Coumou, 2011: Increase of extreme events in a warming world. *Proc. Natl. Acad. Sci.*, **108**, 17 905–17 909, doi: 10.1073/pnas.1101766108.

Räisänen, J., 2002: CO2-induced changes in interannual temperature and precipitation variability in 19 CMIP2 experiments. *J. Climate*, **15**, 2395–2411, doi: 10.1175/1520-0442(2002)015<2395:CICIIT>2.0.CO;2.

Rhines, A. and P. Huybers, 2012: Frequent summer temperature extremes reflect changes in the mean, not the variance. *Proc. Natl. Acad. Sci.*, **110**, E546–E546, doi: 10.1073/pnas.1218748110.

Saffioti, C., E. M. Fischer, S. C. Scherrer, and R. Knutti, 2016: Reconciling observed and modeled temperature and precipitation trends over Europe by adjusting for circulation variability. *Geophys. Res. Lett.*, **43**, 8189–8198, doi: 10.1002/2016GL069802.

Schär, C., P. L. Vidale, D. Lüthi, C. Frei, C. Häberli, M. A. Liniger, and C. Appenzeller, 2004a: The role of increasing temperature variability in European summer heatwaves. *Nature*, **427**, 332–336, doi: 10.1038/nature02230.1.

Schär, C., P. L. Vidale, D. Lüthi, C. Frei, C. Häberli, M. A. Liniger, and C. Appenzeller, 2004b: The role of increasing temperature variability in European summer heatwaves. *Nature*, **427**, 332–336, doi: 10.1038/nature02300.

Scherrer, S. C., M. A. Liniger, and C. Appenzeller, 2008: Distribution changes of seasonal mean temperature in observations and climate change scenarios. *Advances in Global Change Research*, S. B. et al., Ed., Springer, Vol. 33, 251–267, doi: 10.1007/978-1-4020-6766-2_17.

Schneider, E. K. and J. L. Kinter, 1994: An examination of internally generated variability in long climate simulations. *Climate Dynamics*, **10**, 181–204, doi: 10.1007/BF00208987.

Schneider, T., T. Bischoff, and H. Plotka, 2015: Physics of changes in synoptic midlatitude temperature variability. *J. Climate*, **28**, 2312–2331, doi: 10.1175/JCLI-D-14-00632.1.

Screen, J. A., 2013: Influence of Arctic sea ice on European summer precipitation. *Environ. Res. Lett.*, **8**, 044 015, doi: 10.1088/1748-9326/8/4/044015.

Screen, J. A., 2014: Arctic amplification decreases temperature variance in northern mid- to high-latitudes. *Nature Clim. Change*, **4**, 577–582, doi: 10.1038/nclimate2268.

Seneviratne, S. I., D. Lüthi, M. Litschi, and C. Schär, 2006: Land–atmosphere coupling and climate change in Europe. *Nature*, **443**, 205–207, doi: 10.1038/nature05095.

Simolo, C., M. Brunetti, M. Maugeri, and T. Nanni, 2011: Evolution of extreme temperatures in a warming climate. *Geophys. Res. Lett.*, **33**, L16 701, doi: 10.1029/2011GL048437.

Sippel, S., N. Meinshausen, A. Merrifield, F. Lehner, A. G. Pendergrass, E. Fischer, and R. Knutti, 2019: Uncovering the forced climate response from a single ensemble member using statistical learning. *J. Climate*, **32**, 5677–5699, doi: 10.1175/JCLI-D-18-0882.1.

Smoliak, B. V., J. M. Wallace, P. Lin, and Q. Fu, 2015: Dynamical Adjustment of the Northern Hemisphere Surface Air Temperature Field: Methodology and Application to Observations. *J. Climate*, **28**, 1613–1629, doi: 10.1175/JCLI-D-14-00111.1.

Smoliak, B. V., J. M. Wallace, M. T. Stoelinga, and T. P. Mitchell, 2010: Application of partial least squares regression to the diagnosis of year-to-year variations in Pacific Northwest snowpack and Atlantic hurricanes. *Geophys. Res. Lett.*, **37**, L03 801, doi: 10.1029/2009GL041478.

Stouffer, R. J. and R. T. Wetherald, 2007: Changes of variability in response to increasing greenhouse gases. Part I: Temperature. *J. Climate*, **20**, 5455–5467, doi: 10.1175/2007JCLI1384.1.

Sutton, R. T., B. Dong, and J. M. Gregory, 2007: Land/sea warming ratio in response to climate change: IPCC AR4 model results and comparison with observations. *Geophys. Res. Lett.*, **34**, L02 701, doi: 10.1029/2006GL028164.

Tamarin-Brodsky, T., K. Hodges, B. J. Hoskins, and T. G. Shepherd, 2018: A Dynamical Perspective on Atmospheric Temperature Variabilityand Its Response to Climate Change. *J. Climate*, **32**, 1707–1724, doi: 10.1175/JCLI-D-18-0462.1.

Tamarin-Brodsky, T., K. Hodges, B. J. Hoskins, and T. G. Shepherd, 2019: A Dynamical Perspective on Atmospheric Temperature Variability and Its Response to Climate Change. *J. Climate*, **32**, 1707–1724, doi: 10.1175/JCLI-D-18-0462.1.

Tamarin-Brodsky, T., K. I. Hodges, B. J. Hoskins, and T. Shepherd, 2020: Changes in Northern Hemisphere temperature variability shaped by regional warming patterns. *Nat. Geosci.*, **13**, 414–421, doi: 10.1038/s41561-020-0576-3.

Tan, Z., O. Lachmy, and T. A. Shaw, 2019: The sensitivity of the jet stream response to climate change to radiative assumptions. *JAMES*, **11**, 934–956, doi: 10.1029/2018MS001492.

Thompson, D. W. J. and J. M. Wallace, 1998: The arctic oscillation signature in the wintertime geopotential height and temperature fields. *Geophys. Res. Lett.*, **25**, 1297–1300, doi: 10.1029/98GL00950.

Thompson, D. W. J., J. M. Wallace, P. D. Jones, and J. J. Kennedy, 2009: Identifying Signatures of Natural Climate Variability in Time Series of Global-Mean Surface Temperature: Methodology and Insights. *J. Climate*, **22**, 6120–6141, doi: 10.1175/2009JCLI3089.1.

Van den Dool, H. M., 1994: Searching for analogues, how long must we wait? *Tellus*, 314–324.

Van den Dool, H. M., J. Huang, and Y. Fan, 2003: Performance and analysis of the constructed analogue method applied to U.S. soil moisture over 1981-2001. *J. Geophys. Res.*, **108**, 8617, doi: 10.1029/2002JD003114.

Volodin, E. M. and A. Y. Yurova, 2013: Summer temperature standard deviation, skewness and strong positive temperature anomalies in the present day climate and under global warming conditions. *Clim. Dyn.*, **40**, 1387–1398.

von Storch, H. and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research*. Cambridge University Press, 494 pp.

Wallace, J. M., Q. Fu, B. V. Smoliak, P. Lin, and C. M. Johanson, 2012: Simulated versus observed patterns of warming over the extratropical northern hemisphere continents during the cold season. *PNAS*, **109**, 14 337–14 342, doi: 10.1073/pnas.1204875109.

Wallace, J. M., C. Smith, and C. S. Bretherton, 1992: Singular value decomposition and wintertime sea surface temperatures and 500-mb height anomalies. *J. Climate*, **5**, 561–576.

Wallace, J. M., Y. Zhang, and J. A. Renwick, 1995: Dynamic contribution to hemispheric mean temperature trends. *Science*, **270**, 780–783.

Walsh, J., 2014: Intensified warming of the Arctic: Causes and impacts on middle latitudes. *Glob. Planet. Change*, **117**, 52–63.

Wilks, D. S., 2016: The Stippling Shows Statistically Significant Grid Points: How Research Results are Routinely Overstated and Overinterpreted, and What to Do about It. *Bull. Amer. Meteor. Soc.*, **97**, 2263–2273, doi: 10.1175/BAMS-D-15-00267.1.

Wills, R. C. J., D. S. Battisti, K. C. Armour, T. Schneider, and C. Deser, 2020: Pattern recognition methods to separate forced responses from internal variability in climate model ensembles and observations. *J. Climate*, **33**, 8693–8719, doi: 10.1175/JCLI-D-19-0855.1.

Yan, X.-H., T. Boyer, K. Trenberth, T. R. Karl, S.-P. Xie, V. Nieves, K.-K. Tung, and D. Roemmich, 2016: The global warming hiatus: Slowdown or redistribution? *Earth's Future*, **4**, 472–482, doi: 10.1002/2016EF000417.

Yettella, V., J. B. Weiss, J. E. Kay, and A. G. Pendergrass, 2018: An ensemble covariance framework for quantifying forced climate variability and its time of emergence. *J. Climate*, **31**, 4117–4133, doi: 10.1175/JCLI-D-17-0719.1.

Yulaeva, E. and J. M. Wallace, 1994: The signature of ENSO in global temperature and precipitation fields derived from the microwave sounding unit. *J. Climate*, **7**, 1719–1736.

Zeppetello, L. R. V., A. Donohoe, and D. S. Battisti, 2019: Does surface temperature respond to or determine downwelling longwave radiation? *Geophys. Res. Lett.*, **46**, 2781–2789, doi: 10.1029/2019GL082220.

APPENDIX A

Supplemental Results from Chapter 3

A1 SUPPLEMENTARY MATERIAL FOR CHAPTER 3

This Appendix contains supplementary material for Chapter 3. Figure A1 shows the standard deviation on temperature trends in the control run and CESM-LE. While the two panels exhibit visual differences, a F-test finds that they are generally not statistically different. This suggests that internal variability is consistent between the control run and the large ensemble runs.

Figure A2 shows a schematic of the decomposition of temperature trends in a single CESM-LE member. This method was used in Deser et al. (2016) to investigate the internal vs. forced dynamical contribution.

Table A1 compares the PLS regression method from SLP and 950hPa winds across all grid points (not just land). Compared to Table 3.1, we see that the addition of ocean grid points lowers the RMSE values in almost all cases. This is unsurprising, as temperature over ocean regions have less variability than over land. While the results in Chapter 3 focus exclusively on land regions, internal dynamics also play a role on temperature trends over the ocean.

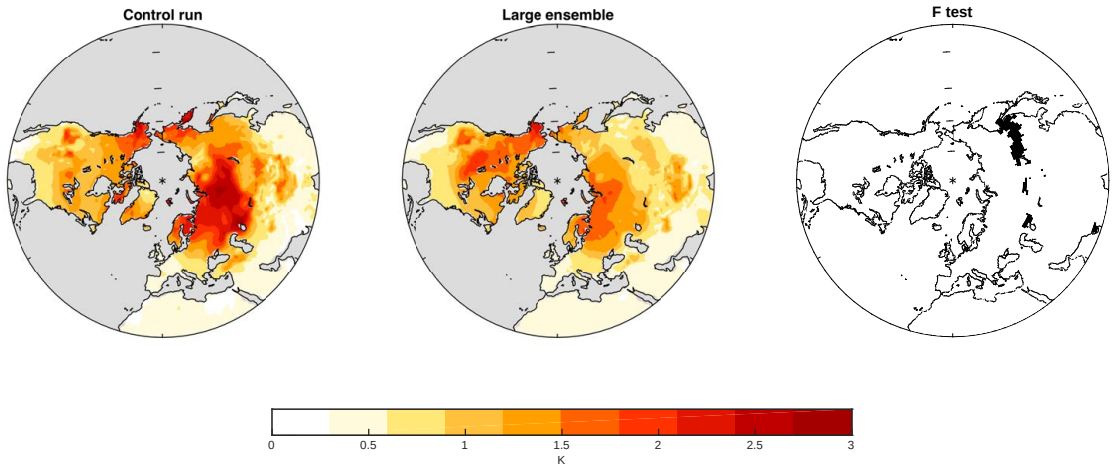

Fig. A1. Comparison of internal variability in the control run and large ensemble. (left) standard deviation on temperature trends in the control run, (middle) standard deviation on temperature trends in the large ensemble, and (right) a F-test between the first two panels, where black signals statistical significance at a field significant level.

	Total	Internal	Forced
Total	Original temperature data from the CESM-LE (single ensemble member)	Internal variability in the ensemble member (Total - Forced)	Ensemble-mean temperatures from the CESM-LE
Dynamic	Dynamically-induced temperatures as calculated by dynamical adjustment	Internal variability due to dynamics (Total - Forced)	Ensemble-mean of all dynamically-induced temperature trends
Thermodynamic	Dynamically-adjusted temperatures (Total - Dynamic)	Internal variability due to thermodynamics (Total - Dynamic)	Ensemble-mean of all dynamically-adjusted temperature trends (Total - Dynamic)

Fig. A2. Schematic showing the decomposition of temperature trends in the CESM-LE into its dynamic and thermodynamic components, and its internal and forced components.

Table A1. RMSE results between the true forced signal (zero in the control run) and dynamically-adjusted temperature trends from each of the 10 50-year periods. The columns show the mean of the RMSE values, the standard deviation, and the single highest RMSE across the 10 ensemble members. Results are calculated from all grid points (land + ocean) in the Northern Hemisphere extratropics (20-90N).

Method	**Mean RMSE**	**Standard Deviation**	**Maximum**
None (unadjusted)	1.0712	0.2620	1.5708
SLP PLS (1 predictor)	0.7695	0.1203	0.9510
SLP PLS (3 predictors)	0.5848	0.0890	0.7080
V950 PLS (1 predictor)	0.7548	0.0931	0.9002
V950 PLS (3 predictors)	0.4967	0.0834	0.6357
U950 + V950 PLS (1 predictor)	0.7027	0.1064	0.8468

APPENDIX B

Supplemental Results from Chapter 4

B1 SURFACE TEMPERATURE AUTOCORRELATIONS IN ERA-INTERIM

The results of Chapter 4 clearly show that there are robust changes in the surface temperature persistence under external forcings in climate change simulations. But how good is the CESM model's representation of SAT persistence? If there are biases in the model, then there may be biases in our results. We compare temperature autocorrelations from the CESM-LE to those derived from ERA-Interim data (Figure B3). 5 day autocorrelations on surface temperature are shown for DJF months, JJA months, and over all calendar days.

We note a few differences in the data and calculations. Daily temperature data from ERA-Interim was obtained over the period 1979-2017 on a spatial grid of 1.5° latitude by 1.5° longitude. Autocorrelations were calculated after detrending the temperature timeseries using linear regression. In contrast, data from CESM-LE is over the period 1970-1999 and on a nominal 1° by 1° resolution. Anomalies are calculated by subtracting the ensemble-mean temperature data (a good estimate of the forced signal), and autocorrelations are ensemble-averaged via the Fisher Z-transformation.

These discrepancies can be seen in Figure B3. We show autocorrelations at a lag of 5 days (as in Figures 4.1 and 4.2) but results are similar for autocorrelations at a lag of 1 and 10 days. The smaller resolution and averaging of 40 ensemble members create much smoother contours in the CESM-LE results (right panels). However, the spatial patterns in the autocorrelation results are very similar between the two datasets. Specifically, the CESM model does a good job capturing the ENSO region, the Atlantic basin, the Southern Ocean, and the Indian Ocean regions. The CESM model does appear to overestimate persistence over tropical ocean regions, however this could be due to the various differences in data and calculations listed above.

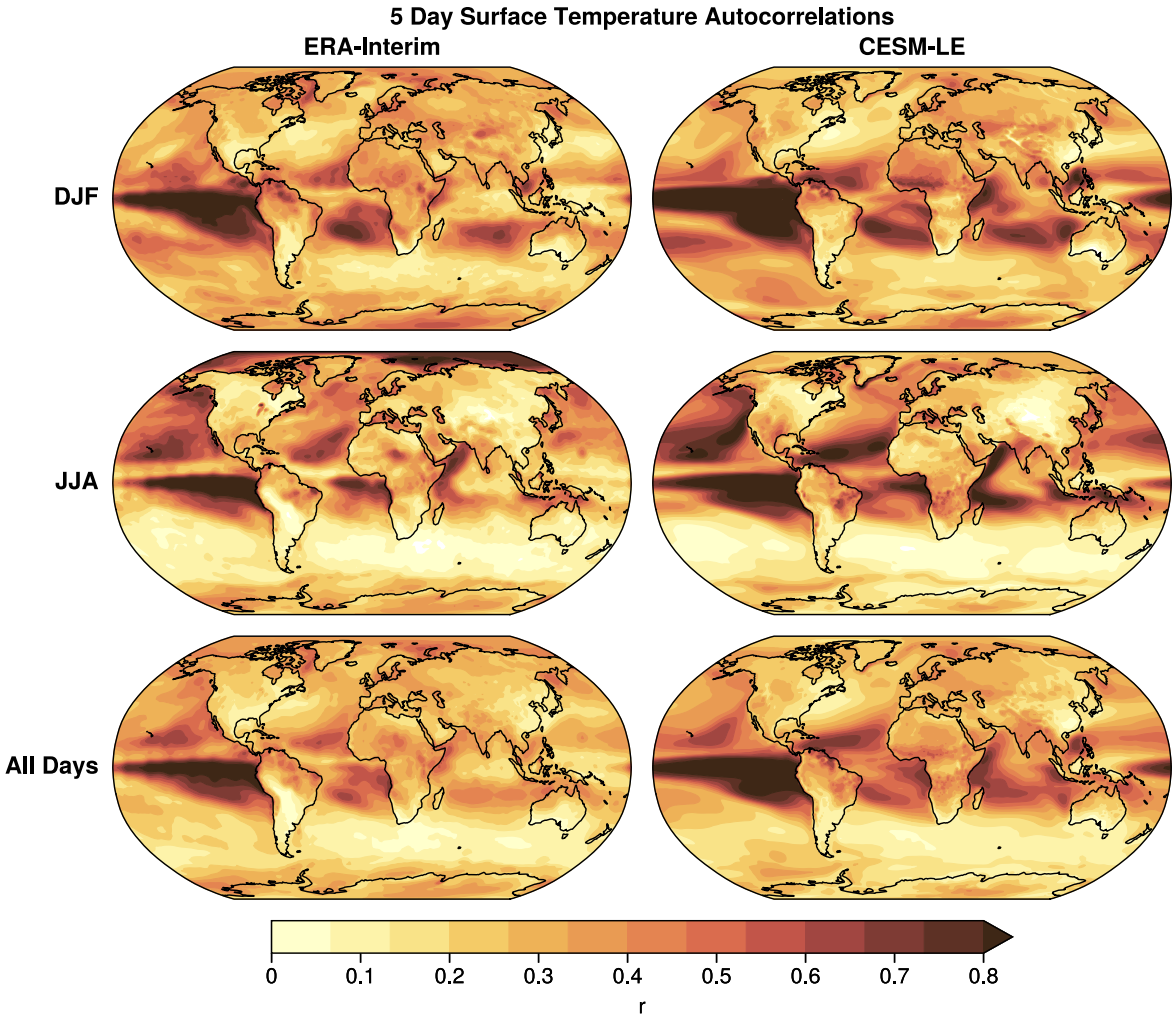

Fig. B3. Comparing surface temperature autocorrelations in the CESM-LE with ERA-Interim. 5 day autocorrelations calculated for (top) DJF, (middle) JJA, and (bottom) all calendar days. ERA-Interim data is over the period 1979-2017. CESM-LE data is over the period 1970-1999 and ensemble averaged.

B2 COMPARING TWO DIFFERENT STATISTICAL SIGNIFICANCE METHODS

In Chapter 4, we apply two different methods to assess the statistical significance of temperature persistence changes in the CESM-LE. The use of a large ensemble allows us to examine significance by directly examining the individual ensemble members, since each ensemble member is completely independent from the others. Here we compare the ensemble agreement method to a traditional statistical significance test. Figure B4 shows grid points that show a statistically significant change in surface temperature autocorrelations at a lag of 5 days between the 1970-1999 and

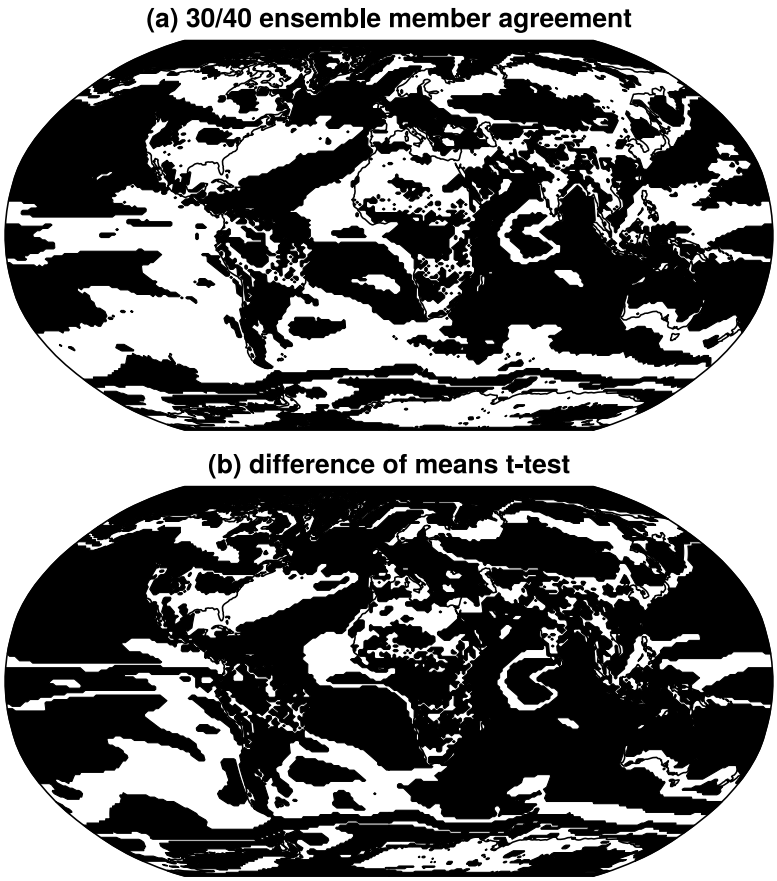

Fig. B4. Grid points exhibiting statistical significance for SAT autocorrelation changes at a lag of 5 days calculated via (a) ensemble agreement, where at least 30 out of 40 ensemble members agree on the sign of the change; (b) a comparison of means t-test with a p-value calculated at the 99% significance level. Black indicates grid points that are calculated to be statistically significant.

2070-2099 periods (the stippling in Figures 4.1 and 4.2) using each method. There are three main takeaways from Figure B4: 1) most of the globe is found to have statistically significant changes in temperature autocorrelation at a lag of 5 days; 2) the two methods show similar spatial patterns in their results and are not very different from one another; 3) the ensemble agreement method is slightly more stringent, as we expected from our probability calculations in 4.2. All statistical significance shown in Chapter 4 figures are defined as grid points that pass both methods.

B3 TEMPERATURE PERSISTENCE CHANGES IN AN AQUAPLANET SIMULATION

A recent study on surface temperature persistence (Pfleiderer et al. (2019)) used a set of AMIP models to investigate the effects of increased surface warming. The authors only examined results on land regions in the Northern Hemisphere, since the SSTs in the climate change runs are fixed at 1.5° and 2°C above the historical runs.

We also investigate SAT persistence changes in an idealized aquaplanet model (the NCAR CAM5.3 global climate model) under perpetual equinox conditions. The simulations are run with historical and 4K forcing, where the SSTs in the forced run are held fixed and uniformly increased by 4K from the historical run values. Figure B5 shows the zonal-mean (calculated via the Fisher Z-transformation) surface air temperature autocorrelations for lags up to 30 days in the historical and +4K runs, and their changes. The strong drop off in autocorrelations after only a few days is due to the model configuration, as the atmosphere is not allowed to feedback into the SSTs. However, the change in autocorrelations with lag is more consistent with lag time, with increases in the poles and tropics, and decreases in the midlatitudes. Figures B6 and B7 focus on the lag one day autocorrelation, where the signal is the strongest.

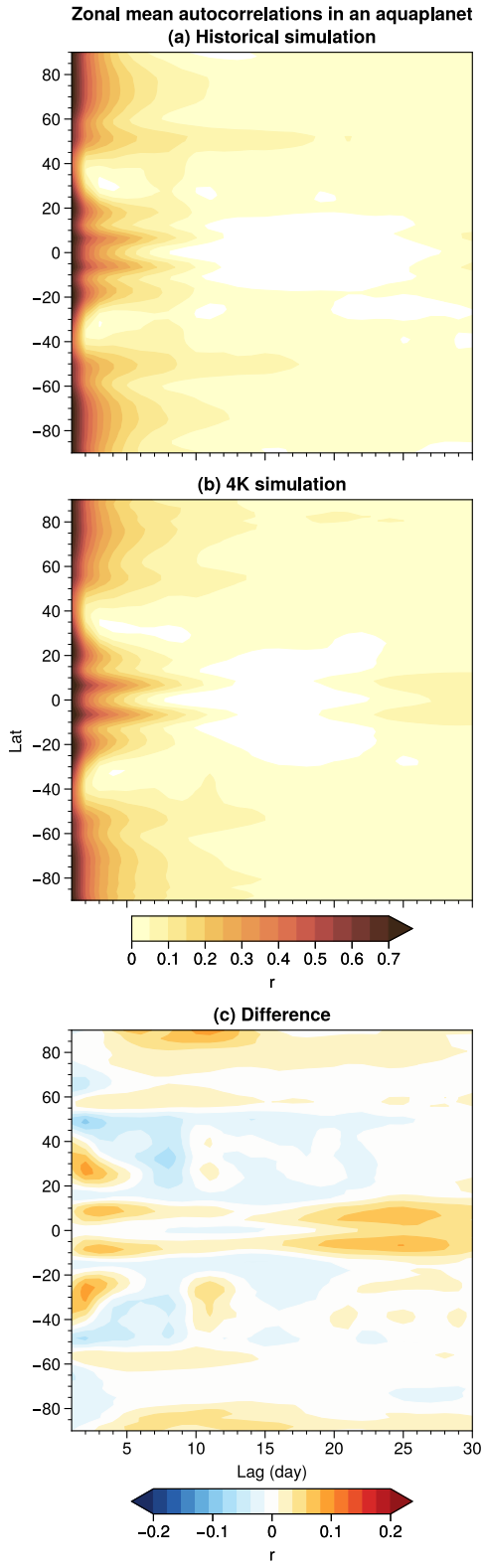

Fig. B5. Zonal-mean surface temperature autocorrelations in an aquaplanet model run for (a) the historical simulation; (b) +4K SST simulation; and (c) the differences in the two simulations. Warm (cool) colors in panel (c) represent an increase in memory under the +4K run.

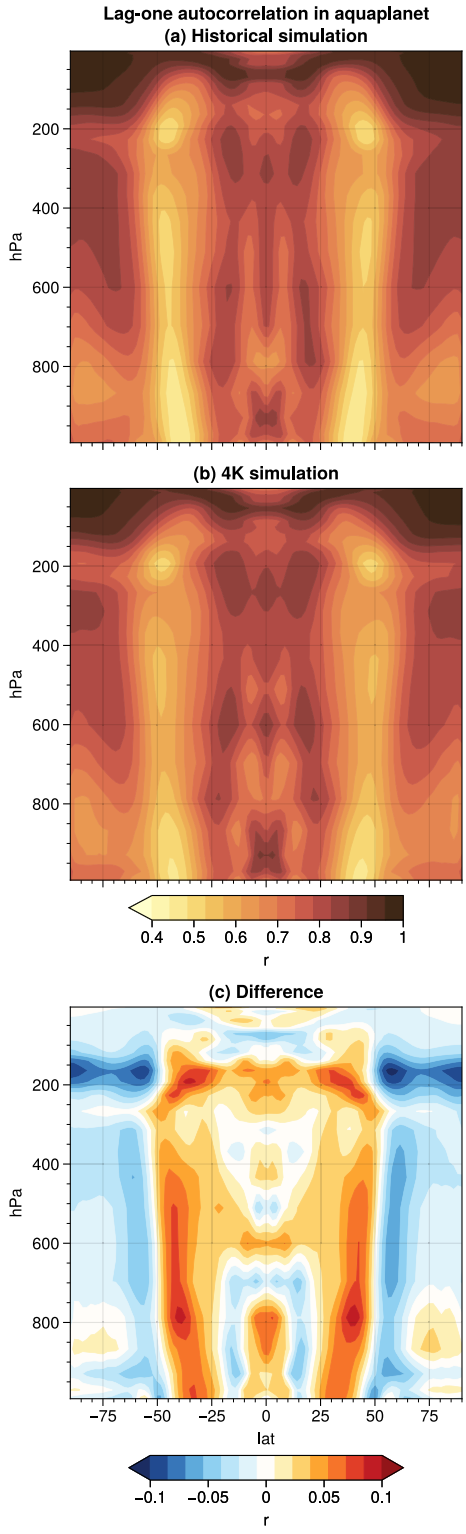

Fig. B6. Zonal mean lag one day surface air temperature autocorrelations in an aquaplanet simulation for (a) the historical simulation; (b) +4K SST simulation; (c) the difference in autocorrelation between the two simulations. Results are shown as a function of latitude and pressure. Warm (cool) colors represent an increase in memory under the +4K simulation.

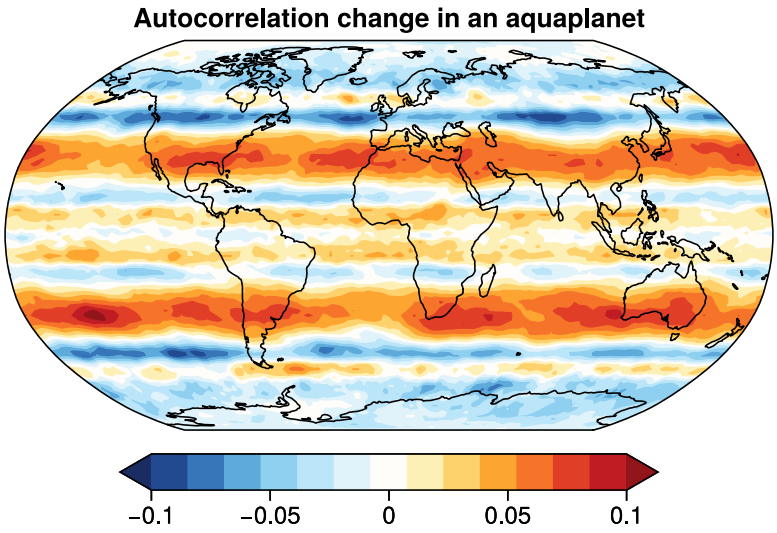

Fig. B7. Change in lag one day surface air temperature autocorrelations in the aquaplanet simulations. Differences are calculated as autocorrelations from the 4K simulation minus those from the historical simulation. Note that the continents are there for location reference only.